Cuk-H02291·64-P04911

AMERICAN JOURNAL OF MATHEMATICS

64

FOUNDED BY THE JOHNS HOPKINS UNIVERSITY

EDITED BY

R. BRAUER
UNIVERSITY OF MICHIGAN

L. M. GRAVES
UNIVERSITY OF CHICAGO

S. EILENBERG
COLUMBIA UNIVERSITY

D. C. LEWIS, JR.
THE JOHNS HOPKINS UNIVERSITY

A. WINTNER
THE JOHNS HOPKINS UNIVERSITY

WITH THE COÖPERATION OF

L. V. AHLFORS
R. ARENS
J. D. HILL
N. H. McCOY
J. L. SYNGE
C. B. ALLENDOERFER
R. BAER
P. HARTMAN
J. J. STOKER
O. SZÁSZ
W. L. CHOW
P. R. HALMOS
H. SAMELSON
R. M. THRALL
H. WEYL

PUBLISHED UNDER THE JOINT AUSPICES OF

THE JOHNS HOPKINS UNIVERSITY
AND
THE AMERICAN MATHEMATICAL SOCIETY

5105
.602

VOLUME LXXII
1950

CALCUTTA UNIVERSITY
CENTRAL LIBRARY

THE JOHNS HOPKINS PRESS
BALTIMORE 18, MARYLAND
U. S. A.

914211

PRINTED IN THE UNITED STATES OF AMERICA
BY J. H. FURST COMPANY, BALTIMORE, MARYLAND

INDEX

PAGE

Baer, Reinhold. Free sums of groups and their generalizations II, 625

Baer, Reinhold. Free sums of groups and their generalizations III, 647

Bers, Lipman. The expansion theorem for sigma-monogenic functions, 705

Blumenthal, Leonard M. Generalized Euclidean space in terms of a quasi inner product, ... 686

Brauer, Alfred and Nathaniel Macon. On the approximation of irrational numbers by the convergents of their continued fractions, II, ... 419

Brauer, Richard. On the zeta-functions of algebraic number fields II, 739

Cesari, Lamberto. On the representation of surfaces, ... 335

Chow, Wei-Liang. Algebraic systems of positive cycles in an algebraic variety, ... 247

Cohen, L. W. and Casper Goffman. On completeness in the sense of Archimedes, ... 747

Cohen, L. W. and Casper Goffman. On completeness and category in uniform space, ... 752

Craig, Homer V. and William T. Guy, Jr. Jacobian extensors, . 229

Davis, Philip. An application of the theory of basic series to theorems of Bernstein-Widder type, ... 787

Doss, Raouf. Some theorems on almost periodic functions, . . 81

Errata, ... 624

Everett, C. J. Note on a result of L. Fuchs on ordered groups, . 216

Fort, M. K., Jr. Essential and non essential fixed points, . . . 315

Foster, Alfred L. On n-ality theories in rings and their logical algebras, including tri-ality principle in three valued logics, 101

Frucht, Robert. On the construction of partially ordered system with a given group of automorphisms, ... 195

Fuchs, Ladislas. On the extension of the partial order of groups, 191

Goodstein, R. L. The Gauss test for relative convergence, . . . 217

Goffman, Casper. See Cohen and Goffman, p. 747.

Goffman, Casper. See Cohen and Goffman, p. 752.

PAGE

Grosswald, Emil. On the structure of some subgroups of the modular group, ... 809

Guy, William T., Jr. See Craig and Guy, page 229.

Halmos, Paul R. and Herbert E. Vaughan. The marriage problem, 214

Hartman, Philip and Aurel Wintner. On linear difference equations of second order, ... 124

Hartman, Philip and Aurel Wintner. On the derivatives of the solutions of one-dimensional wave equations, ... 148

Hartman, Philip and Aurel Wintner. On non-linear differential equations of first order, ... 347

Hartman, Philip and Aurel Wintner. On the spectra of Toeplitz's matrices, ... 359

Hartman, Philip and Aurel Wintner. On the solutions of the equation of heat conduction, ... 367

Hartman, Philip and Aurel Wintner. On the essential spectra of singular eigenvalue problems, ... 545

Hartman, Philip and Aurel Wintner. On the embedding problem in differential geometry, ... 553

Hartman, Philip. On the local uniqueness of geodesics, ... 723

Hartman, Philip and Aurel Wintner. On the fundamental equations of differential geometry, ... 757

Hartman, Philip and Aurel Wintner. Lamé coordinates in Hilbert space, ... 775

Hartman, Philip and Calvin R. Putnam. The gaps in the essential spectra of wave equations, ... 849

Hill, J. D. Summability methods weaker than convergence, . . . 621

Hirschman, I. I., Jr. On the behaviour of Fourier transforms at infinity and on quasi-analytic classes of functions, . . . 200

Hirschman, I. I., Jr. On the distributions of the zeros of functions belonging to certain quasi-analytic classes, ... 396

Hirschman, I. I., Jr. On non-uniformly quasi-analytic functions, . 863

Hochschild, G. Lie algebras and differentiations in rings of power series, ... 58

Hölder, Ernst. Zusätzliche Stabilitätsbetrachtung betreffend "Die Symmetrischen Periodischen Bahnen des Restringierten Dreikörperproblems in der Nachbarschaft eines kritischen Keplerkreises," 157

Jackson, S. B. Geodesic vertices on surfaces of constant curvature, . 161

PAGE

Kaplan, Wilfred. Dynamical systems with indeterminacy, . . 573

Kelly, John B. A closed set of algebraic integers, . . . 565

Kodaira, Kunihiko. On ordinary differential equations of any even order and the corresponding eigenfunction expansions, . 502

Kuiper, N. H. On linear familes of involutions, ... 425

Loo, Ching-Tsün. On the uniform Cesáro summability of certain special trigonometrical series, ... 129

Macon, Nathaniel. See Brauer and Macon, page 419.

Magnus, Wilhelm. On the spectrum of Hilbert's matrix, . . 699

Malin, Henry. See Philips and Malin, page 407.

Martin, Monroe H. Steady, rotational, plane flow of a gas, . . 465

Neumann, Hanna. Generalized free sums of cyclical groups, . . 671

Peck, L. G. On cube-coverings of n-space, ... 720

Philips, R. S. and Henry Malin. Bessel function approximations, 407

Putnam, C. R. On isolated eigenfunctions associated with bounded potentials, ... 135

Rickart, C. E. Isomorphic groups of linear transformations, . . 451

Scheid, Francis. On the asymptotic shape of the cavity behind an axially symmetric nose moving through an ideal fluid, . 485

Sheffer, I. M. On the solution of sum-equations, ... 835

Sherman, S. On denumerably independent families of Borel fields, . 612

Sips, Robert. Quelques propriétés extrémales du cercle et de la sphère, 595

Smiley, Malcolm F. Application of a radical of Brown and McCoy to non-associative rings, ... 93

Tornheim, Leonard. Harmonic double series, ... 303

Vaughan, Herbert E. See Halmos and Vaughan, page 214.

Vermes, P. Certain classes of series to series transformation matrices, 615

Wang, Shianghaw. On the commutator group of a simple algebra, 323

Ward, Morgan. Arithmetical properties of the elliptic polynomials arising from the real multiplication of the Jacobi functions, . 284

Whitehead, J. H. C. Simple homotopy types, ... 1

Wilkins, J. Ernest, Jr. The general term of the generalized Schlömilch series, ... 187

Wing, G. Milton. The mean convergence of orthogonal series, . . 792

PAGE

WINTNER, AUREL. See Hartman and Wintner, page 124.

WINTNER, AUREL. See Hartman and Wintner, page 347.

WINTNER, AUREL. See Hartman and Wintner, page 359.

WINTNER, AUREL. See Hartman and Wintner, page 367.

WINTNER, AUREL. See Hartman and Wintner, page 545.

WINTNER, AUREL. See Hartman and Wintner, page 553.

WINTNER, AUREL. See Hartman and Wintner, page 757.

WINTNER, AUREL. See Hartman and Wintner, page 775.

WINTNER, AUREL. On the existence of Laplace solutions for linear differential equations of second order, ... 442

WINTNER, AUREL. On the Hölder restrictions in the theory of partial differential equations, ... 731

WOLFSON, KENNETH G. On the spectrum of a boundary value problem with two singular end-points, ... 713

CENTRAL LIBRARY

SIMPLE HOMOTOPY TYPES.*

By J. H. C. Whitehead.

---

1. Introduction. This is a sequel to two papers[1] entitled "Combinatorial Homotopy," Parts (I) and (II). It deals with what I have previously called the "nucleus," but which will now be called the *simple homotopy type* of a complex. It is closely related to parts of [1] and [3] but the treatment is so different that we shall start again from the beginning.

Let $\{K\}$ be the class of all (cell) complexes,[2] as defined in CH (I), which are of the same homotopy type as a given complex K. Let $K' \equiv K$ (i. e. $K' \varepsilon \{K\}$) and let $\bar{\phi}: K \equiv K'$ be the class of maps which are homotopic to a given homotopy equivalence, $\phi: K \equiv K'$. If $\phi': K' \equiv K''$, we define $\bar{\phi}'\bar{\phi}$ by

$$\bar{\phi}'\bar{\phi} = \overline{\phi'\phi}: K \equiv K''.$$

It is easily verified that the classes ϕ, with this multiplication, form a groupoid,[3] G, whose unit elements are the classes $\bar{1}: K' \equiv K'$, for every $K' \varepsilon \{K\}$, where $1: K' \rightarrow K'$ is the identical map. Our plan is to analyse this groupoid in algebraic terms.

First consider the group, $G_K \subset G$, which consists of the classes $\bar{\phi}: K \equiv K$. We define an additive Abelian group, T, which depends only on $\pi_1(K)$. The group T admits G_K as a group of operators and we shall define a crossed homomorphism $\tau: G_K \rightarrow \mathrm{T}$. We call $\tau(g)$ the torsion of a given element $g \varepsilon G_K$. If $\phi: K \equiv K'$, where $K' \neq K$, we define a class of elements $\tau(\phi) \subset \mathrm{T}$, which we call the *torsion* of ϕ. We describe ϕ as a *simple* (*homotopy*) *equivalence* if, and only if, $\tau(\phi) = 0$. We say that K and K' are of the same *simple homotopy type*, and shall write $K \equiv K'\ (\Sigma)$, if, and only if, there is a simple equivalence $\phi: K \equiv K'$. It will follow from the

---

* Received January 18, 1949.

[1] *Bulletin of the American Mathematical Society*, vol. 55 (1949), pp. 213-45 and 453-96. These papers will be referred to as CH (I) and CH (II).

[2] Until the final section we assume that any given complex is finite and connected. We also assume that the points in our complexes are taken from some aggregate, σ, which is given in advance. The power of σ shall exceed that of the continuum, so that it is not exhausted by any one (finite) complex, and the points in Hilbert space shall be included in σ.

[3] See [6], p. 132.

definition of $\tau(\phi)$ that $K \equiv K' (\Sigma)$ is an equivalence relation. We then prove that $K \equiv K' (\Sigma)$ if, and only if, K can be transformed into K' by a "formal deformation," which is defined in much the same way as in [1]. Thus the elementary transformations, or "moves," do not appear in the definition of simple equivalence but in a theorem which is analogous to Tietze's theorem[4] on discrete groups. Similarly it is proved that two complexes are of the same n-type if, and only if, they can be interchanged by elementary transformations of the sort used in [1] to define the "n-group."

It was proved in [3] that the Reidemeister-Franz torsion,[5] when defined, is an invariant of the simple homotopy type. Using this fact, examples were given of complexes, which are of the same homotopy type but not of the same simple homotopy type. However, if $\mathrm{T} = 0$, then $K \equiv K' (\Sigma)$ if $K \equiv K'$. It will be obvious that this is so if $\pi_1(K) = 1$. It follows from Theorems 14, 15 in [11] that $\mathrm{T} = 0$ if $\pi_1(K)$ is of order 2, 3, 4 or cyclic infinite.

It is an open question whether or not the simple homotopy type is a topological invariant. However we shall prove that it is a combinatorial invariant in the following sense. If K' is a sub-division of K, then the identical map $K \rightarrow K'$ is a simple equivalence.[6] Any differentiable manifold has a "preferred" class of triangulations,[7] any two of which are combinatorially equivalent in the sense of Newman. Also any analytic variety has a preferred class of triangulations,[7] any two of which have a common subdivision. Therefore the simple homotopy type has an invariant status in differential and alegbraic geometry and in the study of analytic varieties.

2. The group T. Let R be a ring with a unit element 1. Eventually R will be the group ring[8] of $\pi_1(K)$ but here we only assume that, if A is a free R-module of (finite) rank n, then any free R-module, which is isomorphic[9] to A, also has rank n. This condition is equivalent to the condition that every regular R-matrix (i. e. one with elements in R and a 2-sided inverse) is square. Hence it is satisfied if there is a homomorphism, other than $R \to 0$, of R into a division ring, D. For such a homomorphism carries a regular R-matrix into a regular D-matrix, which is necessarily square. If R is the group ring of a group, Γ, then $\Gamma \to 1$ defines a homomorphism of R into the rational field. Therefore the condition of rank invariance is satisfied.

---

[4] See [7], p. 46.

[5] See [8], [9] and p. 1209 of [3]. In Section **12** below it is shown, in the case of Lens space, how this is related to our torsion. See also [10].

[6] This may turn out to be a wider definition, even for simplicial complexes, than the one based on Newman's "moves," or on recti-linear sub-divisions (see [12], [13]). For example, we do not enquire whether or not the vertex scheme of a given "curvilinear" triangulation of an n-simplex is a formal n-element, as defined by Newman.

[7] See [2] and [14].

[8] By the group ring of a group, Γ, we shall always mean the integral group ring, in which the additive group is the ordinary free Abelian group, which is freely generated by the elements of Γ.

[9] A module will always mean a free R-module and, unless the contrary is stated, a homomorphism will always mean an operator homomorphism.

Let M be the module, of infinite rank, whose elements are the infinite sequences $(r_1, r_2, \cdots)$ $(r_i \varepsilon R)$, in which all but a finite number of $r_1, r_2, \cdots$ are zero. The elements in R will operate on M from the left.[10] Thus an operator, $r \varepsilon R$, transforms m into rm, where

$$m = (r_1, r_2, \cdots), \quad rm = (rr_1, rr_2, \cdots).$$

Let $m_i \varepsilon M$ be the basis element which is given by $r_i = 1$, $r_j = 0$ if $j \neq i$. Let $M^n \subset M$ be the module generated[11] by $(m_1, \cdots, m_n)$ and M_n the one generated by $(m_{n+1}, m_{n+2}, \cdots)$, where $n \geqq 0$ and $M^0 = 0$. Then M is the direct sum $M = M^n + M_n$ and a given element in M is in M^n for some value of n. We shall describe an endomorphism, $f: M \to M$, as *admissible* if, and only if, $fm_i = m_i$ for all sufficiently large values of i. If $f, g: M \to M$ are admissible endomorphisms[12] so, obviously, is $fg: M \to M$ and if $f: M \to M$ is an admissible automorphism so is f^{-1}. Therefore the admissible automorphims form a group, $\mathfrak{a}$.

Let $f: M \to M$ be an (admissible) endomorphism and let $fm_j = m_j$ if $j > p$. Let n_i be such that $fm_i \varepsilon M^{n_i}$ $(i = 1, \cdots, p)$ and let $n \geqq \mathrm{Max}(n_i, p)$. Then $fm_i \varepsilon M^n$ for $i = 1, \cdots, n$ and $fm_j = m_j$ if $j > n$. Therefore $fM^n \subset M^n$, $fm = m$ if $m \varepsilon M_n$. We shall write $f = (f)^n: M \to M$, and f^n will denote the endomorphism, $f^n: M^n \to M^n$, which is induced by f. That is to say, $f^n m = fm$ if $m \varepsilon M^n$. Notice that $(f)^n = (f)^q$ if $q > n$. Therefore, if $f_i: M \to M$ is any finite set of endomorphisms, we may take $f_i = (f_i)^n$, for any value of n which is sufficiently large to be the same for each i. Notice also that any endomorphism $f': M^n \to M^n$ can be extended to a unique endomorphism, $(f)^n: M \to M$, such that $f^n = f'$. Obviously f^n is an automorphism if, and only if, $f \varepsilon \mathfrak{a}$.

Let $f = (f)^n$ be given by

$$fm_i = \sum_{j=1}^{\infty} f_{ij} m_j \qquad (f_{ij} \varepsilon R). \tag{2.1}$$

---

[10] This has the disadvantage indicated by (2. 3) below. But the convention $m \to mr$ would be inconvenient in the geometrical application.

[11] I. e. generated with the help of the operators in R.

[12] Unless the contrary is stated it is to be assumed that any given endomorphism of M is admissible.

Then the matrix $\boldsymbol{f} = [f_{ij}]$ is of the form

$$\boldsymbol{f} = \begin{bmatrix} \boldsymbol{f}^n & 0 \\ 0 & \mathbf{1}_\infty \end{bmatrix} \tag{2.2}$$

where $\boldsymbol{f}^n$ is the matrix of $f^n : M^n \rightarrow M^n$ and $\mathbf{1}_\infty$ is the infinite unit matrix. Let $g : M \rightarrow M$ be given by

$$gm_i = \sum_j g_{ij} m_j.$$

Since $fr = rf$, where $r \varepsilon R$ is any operator, we have

$$fgm_i = \sum_j g_{ij} f m_j = \sum_{j,k} g_{ij} f_{jk} m_k. \tag{2.3}$$

Therefore $fg : M \rightarrow M$ corresponds to the matrix $\boldsymbol{gf}$.

Let $g : M \rightarrow M$ be given by

$$gm_i = m_i + rm_j, \qquad gm_k = m_k \qquad (j, k \neq i; r \varepsilon R). \tag{2.4}$$

Then g has an inverse, which is given by (2.4), with r replaced by $-r$. It is therefore an (admissible) automorphism. Let $\Sigma_1 \subset \mathfrak{A}$ be the group generated by all such automorphisms, for all values of i, j, r.

Let A and B be the modules generated by disjoint sub-sets, $m_{i_1}, \cdots, m_{i_p}$ and $m_{j_1}, \cdots, m_{j_q}$, of the basis elements $m_1, m_2, \cdots,$. Let $h : A \rightarrow B$ be an arbitrary homomorphism and let $g : M \rightarrow M$ be given by

$$g(a + b) = a + (ha + b), \qquad gm_l = m_l, \tag{2.5}$$

where $a \varepsilon A$, $b \varepsilon B$ and $l \neq i_\rho$ or j_σ. Then g is the resultant of the homomorphisms

$$m_{i_\rho} \rightarrow m_{i_\rho} + h_{\rho\sigma} m_{j_\sigma}, \qquad m_k \rightarrow m_k \qquad (k \neq i_\rho),$$

where $hm_{i_\rho} = h_{\rho 1} m_{j_1} + \cdots + h_{\rho q} m_{j_q}$. These are of the form (2.4), whence $g \varepsilon \Sigma_1$.

THEOREM 1. *Σ_1 is an invariant sub-group of $\mathfrak{A}$ and $\mathfrak{A}/\Sigma_1$ is Abelian.*

Let $f, f' \subset \mathfrak{A}$. We shall write $f \equiv f'$ if, and only if, $f = gf'g'$, where $g, g' \subset \Sigma_1$. This is obviously an equivalence relation. Assume that $ff' \equiv f'f$ for every pair $f, f' \subset \mathfrak{A}$. Let $g \varepsilon \Sigma_1$, and let $f' = gf^{-1}$. Then

$$fgf^{-1} \equiv gf^{-1}f = g.$$

Therefore $fgf^{-1} \varepsilon \Sigma_1$, whence Σ_1 is invariant in $\mathfrak{A}$. Also $\mathfrak{A}/\Sigma_1$ is Abelian since $ff' \equiv f'f$ for every pair $f, f' \subset \mathfrak{A}$.

We proceed to prove that $ff' \equiv f'f$. Let $A = M^p$, let B be the module generated by $m_{p+1}, \cdots, m_{2p}$ and let g be given by (2.5). Then $g = (g)^{2p}$ and

$$g^{2p} = \begin{bmatrix} \mathbf{1}_p & \boldsymbol{h} \\ 0 & \mathbf{1}_p \end{bmatrix},$$

where $\boldsymbol{h} = [h_{\rho\sigma}]$ and $\mathbf{1}_p$ is the unit matrix of order p. Let $f = (f)^{2p} \varepsilon \mathfrak{A}$ and let

$$\boldsymbol{f}^{2p} = \begin{bmatrix} \boldsymbol{f}_{11} & \boldsymbol{f}_{12} \\ \boldsymbol{f}_{21} & \boldsymbol{f}_{22} \end{bmatrix},$$

where $\boldsymbol{f}_{11}$, $\boldsymbol{f}_{22}$ are square matrices of order p. Let $\boldsymbol{f}'^{2p}$ and $\boldsymbol{f}'_{\lambda\mu}$ $(\lambda, \mu = 1, 2)$ be similarly defined in terms of $f' = (f')^{2p}$. We shall write $\boldsymbol{f}^{2p} \equiv \boldsymbol{f}'^{2p}$ if, and only if, $f \equiv f'$. Then

$$(2.6) \qquad \boldsymbol{f}^{2p} \equiv \boldsymbol{f}^{2p} g^{2p} = \begin{bmatrix} \boldsymbol{f}_{11} & \boldsymbol{f}_{11}\boldsymbol{h} + \boldsymbol{f}_{12} \\ \boldsymbol{f}_{21} & \boldsymbol{f}_{21}\boldsymbol{h} + \boldsymbol{f}_{22} \end{bmatrix}.$$

Similarly a right hand multiple of the second column may be added to the first. Also $\boldsymbol{f}^{2p} \equiv g^{2p}\boldsymbol{f}^{2p}$, and $g^{2p}\boldsymbol{f}^{2p}$ is obtained from $\boldsymbol{f}^{2p}$ by a similar operation on the rows.

Let $f, f' \subset \mathfrak{A}$ be given and let p be so large that

$$f = (f)^p = (f)^{2p}, \qquad f' = (f')^p = (f')^{2p}.$$

Let $\boldsymbol{r} = \boldsymbol{f}^p$, $\boldsymbol{r}' = \boldsymbol{f}'^p$. Then $\boldsymbol{r}$, $\boldsymbol{r}'$ are regular matrices. Therefore, beginning with (2.6), with $\boldsymbol{h} = \boldsymbol{r}^{-1}$ and $\boldsymbol{f}$ replaced by $\boldsymbol{f}'\boldsymbol{f}$, we have

$$\begin{aligned} \boldsymbol{f}'^{2p}\boldsymbol{f}^{2p} &= \begin{bmatrix} \boldsymbol{r}'\boldsymbol{r} & 0 \\ 0 & \mathbf{1}_p \end{bmatrix} \equiv \begin{bmatrix} \boldsymbol{r}'\boldsymbol{r} & \boldsymbol{r}' \\ 0 & \mathbf{1}_p \end{bmatrix} \equiv \begin{bmatrix} 0 & \boldsymbol{r}' \\ -\boldsymbol{r} & \mathbf{1}_p \end{bmatrix} \\ &\equiv \begin{bmatrix} 0 & \boldsymbol{r}' \\ -\boldsymbol{r} & 0 \end{bmatrix} \equiv \begin{bmatrix} \boldsymbol{r} & \boldsymbol{r}' \\ -\boldsymbol{r} & 0 \end{bmatrix} \equiv \begin{bmatrix} \boldsymbol{r} & \boldsymbol{r}' \\ 0 & \boldsymbol{r}' \end{bmatrix} \equiv \begin{bmatrix} \boldsymbol{r} & 0 \\ 0 & \boldsymbol{r}' \end{bmatrix}. \end{aligned}$$

Similarly

$$\begin{bmatrix} 0 & \boldsymbol{r}' \\ -\boldsymbol{r} & 0 \end{bmatrix} \equiv \begin{bmatrix} \boldsymbol{r}' & 0 \\ 0 & \boldsymbol{r} \end{bmatrix} \equiv \boldsymbol{f}^{2p}\boldsymbol{f}'^{2p}.$$

Therefore $\boldsymbol{f}\boldsymbol{f}' \equiv \boldsymbol{f}'\boldsymbol{f}$ and the theorem is proved.

Since $\mathfrak{A}/\Sigma_1$ is Abelian it follows that $\mathfrak{A}^c \subset \Sigma_1$, where $\mathfrak{A}^c$ is the commutator sub-group of $\mathfrak{A}$. Therefore we have the corollary:

COROLLARY. *If $\Sigma \subset \mathfrak{A}$ is any sub-group, which contains Σ_1, then Σ is invariant and $\mathfrak{A}/\Sigma$ is Abelian.*

The totality of automorphisms $(f)^n \varepsilon \mathfrak{A}$, for a fixed value of n, is obviously a sub-group, $(\mathfrak{A})^n \subset \mathfrak{A}$. It follows from Theorem 1 that

$(\Sigma_1)^n = \Sigma_1 \cap (\mathfrak{A})^n$ is an invariant sub-group[13] of $(\mathfrak{A})^n$ and that $(\mathfrak{A})^n/(\Sigma_1)^n$ is Abelian. Let $\mathfrak{A}^n$ be the group of (operator) automorphisms, $f^n: M^n \to M^n$, and let $\phi: (\mathfrak{A})^n \to \mathfrak{A}^n$ be given by $\phi(f)^n = f^n$. Then ϕ is obviously an isomorphism.[14] It follows from the invariance of $(\Sigma_1)^n$ in $(\mathfrak{A})^n$ that $\Sigma_1{}^n = \phi(\Sigma_1)^n$ is invariant in $\mathfrak{A}^n$ and that $\Sigma_1{}^n$ is independent of the particular isomorphism $\phi: (\mathfrak{A})^n \approx \mathfrak{A}^n$. Also $\mathfrak{A}^n/\Sigma_1{}^n$ is Abelian.

Let Λ be a sub-group of the multiplicative group of regular elements in R (that is, elements with two-sided inverses), which contains both ± 1. Let $g: M \to M$ be given by

$$gm_i = \lambda m_i + rm_j, \qquad gm_k = m_k \qquad (j, k \neq i), \tag{2.7}$$

where $\lambda \varepsilon \Lambda$, $r \varepsilon R$. Then $g \varepsilon \mathfrak{A}$ and g^{-1} is given by (2.7) with λ, r replaced by λ^{-1}, $-\lambda^{-1}r$. Let Σ_Λ be the sub-group of $\mathfrak{A}$, which is generated by all automorphisms of the form (2.7), for every choice of i, j, λ and r. Clearly $\Sigma_1 \subset \Sigma_\Lambda$. Therefore Σ_Λ is invariant and $\mathrm{T} = \mathfrak{A}/\Sigma_\Lambda$ is Abelian. We shall keep Λ fixed and shall write Σ, T for Σ_Λ, T_Λ. The elements of Σ will be called *simple automorphisms.* We shall write T additively and $\tau(f) \varepsilon \mathrm{T}$ will denote the co-set containing a given $f \varepsilon \mathfrak{A}$.

Our "torsion" will be defined in terms of T. An element of torsion will correspond to an isomorphism of one module, of finite rank, onto another. In order to classify such isomorphisms in term of T we need a standard class, which have "zero torsion." We therefore proceed to define a class of "basic modules" in M, which are related by a standard class of automorphisms, called permutations.

By a *basic module,* $A \subset M$, we shall mean the one generated by $m_{i_1}, \cdots, m_{i_p}$, for any (distinct) values of $i_1, \cdots, i_p$. We shall call $(m_{i_1}, \cdots, m_{i_p})$ *the* basis of A. We allow $p = 0$, in which case the set $(m_{i_1}, \cdots, m_{i_p})$ is empty and $A = M^0$. Let $p \geqq 0$ and let M_A be the module generated by the remaining basic elements, $m_j \neq m_{i_p}$, of M. Then M is the direct sum $M = A + M_A$. Let B be a basic module and let $(m_{j_1}, \cdots, m_{j_q})$ be its basis. We shall only allow ourselves to form the direct sum $A + B = B + A$, if $A \cap B = 0$. In this case $A + B$ will be the basic module, whose basis is

$$(m_{i_1}, \cdots, m_{i_p}, m_{j_1}, \cdots, m_{j_q}),$$

---

[13] The example II, on p. 1233 of [3] shows that $(\Sigma_1)^n$ may be a larger group than the one which is generated by transformations of the form (2.4), with $i, j \leqq n$. I see no reason to suppose that the latter is necessarily an invariant sub-group of $(a)^n$.

[14] An isomorphism, without qualification, will always mean an isomorphism onto.

not the set of all pairs (a, b), with $a \varepsilon A$, $b \varepsilon B$. Let C be a given basic module. Then $C = A + B$ will always mean that A, B are basic modules, with disjoint bases, of which C is the direct sum.

Let $i_1, \cdots, i_n$ be any permutation of $1, \cdots, n$, for any $n \geqq 1$. Let $P : M \rightarrow M$ be the automorphism, which is given by

$$Pm_j = m_{i_j}, \qquad Pm_k = m_k \qquad (j = 1, \cdots, n; k > n).$$

We shall call P a *permutation.* It follows from (2. 7), with $\lambda, r = \pm 1$, that the transformations

$$(m_i, m_j) \rightarrow (-m_i + m_j, m_j) \rightarrow (-m_i + m_j, m_i) \rightarrow (m_j, m_i)$$

determine simple automorphisms. Therefore $P \varepsilon \Sigma$. Let A, B be basic modules of the same rank and let n be so large that the bases of A, B are both contained in M^n. Then there is obviously a permutation, $P = (P)^n$, such that $PA = B$. The totality of permutations is obviously a sub-group of $\mathfrak{A}$.

Let $\alpha : A \approx A'$, where A, A' are basic modules. Since A and A' have the same rank, accoording to our condition on R, there is a permutation, P, such that $PA' = A$. Let $f : M \rightarrow M$ be given by

$$f(a + m) = P\alpha a + m \qquad (a \varepsilon A, m \varepsilon M_A), \tag{2. 8}$$

and let $\tau(\alpha) = \tau(f)$. Let P' be any other permutation such that $P'A' = A$ and let f' be defined by (2. 8), with P replaced by P'. Since $P'P^{-1}A = A$ the permutation $P'P^{-1}$ permutes the basis elements of A among themselves. Therefore $P'' : M \rightarrow M$, given by

$$P''(a + m) = P'P^{-1}a + m \qquad (m \varepsilon M_A),$$

is a permutation. Since $P\alpha A = A$, it follows from (2. 8) that $f' = P''f$. Therefore

$$\tau(f') = \tau(P'') + \tau(f) = \tau(f).$$

Therefore $\tau(\alpha)$ does not depend on the choice of P. We shall describe α as a *simple isomorphism,* and shall write $\alpha : A \approx A'$ (Σ), if, and only if, $\tau(\alpha) = 0$. It follows from (2. 8) that $\tau(\alpha) = 0$ if $\alpha = 1 : A \approx A$. In particular $\tau(\alpha) = 0$ if $A = A' = M^0$.

Let α, P, f, mean the same as in (2. 8), let $\alpha' : A' \approx A''$ and let P' be a permutation such that $P'A'' = A'$. Then $\tau(\alpha') = \tau(f')$, $\tau(\alpha'\alpha) = \tau(f'')$, where f' and f'' are given by (2. 8) with α, P replaced by α', P' and by $\alpha'\alpha$, PP'. Clearly $P^{-1}M_A = M_{A'}$. Therefore

$$Pf'P^{-1}f(a+m) = Pf'P^{-1}(P\alpha a + m) = Pf'(\alpha a + P^{-1}m)$$
$$= P(P'\alpha'\alpha a + P^{-1}m) = PP'\alpha'\alpha a + m = f''(a+m).$$

Therefore

$$\tau(\alpha'\alpha) = \tau(Pf'P^{-1}f) = \tau(f') + \tau(f) = \tau(\alpha') + \tau(\alpha).$$

Since $\tau(1) = 0$ it follows that $\tau(\alpha^{-1}) = -\tau(\alpha)$.

Let $\alpha : A \approx A'$ be the isomorphism induced by a permutation, $P' : M \to M$, such that $P'A = A'$. Then f, given by (2.8), is a permutation. Therefore $\tau(\alpha) = 0$.

Let A, B and A', B' be two pairs of basic modules such that $A \cap B = A' \cap B' = 0$. Let $\gamma : A + B \to A' + B'$ be a homomorphism such that $\gamma B \subset B'$. Then $\gamma(a+b) = \alpha a + (ha + \beta b)$, where $\alpha a \varepsilon A', ha, \beta b \subset B'$. It is easily verified that $\alpha : A \to A'$, $h : A \to B'$, $\beta : B \to B'$ are homomorphisms.

THEOREM 2. *If either*:

(i) γ *is an isomorphism*[14] *and either* α *is an isomorphism into or* β *is onto, or if*

(ii) α, β *are isomorphisms, then* α, β, γ *are all isomorphisms and*

$$\tau(\gamma) = \tau(\alpha) + \tau(\beta).$$

Let $\gamma : A + B \approx A' + B'$. If $\beta b = 0$, then $\gamma b = \beta b = 0$, whence $b = 0$. Therefore β is an isomorphism into. Let $a' \varepsilon A'$ be given. Then $\alpha a + (ha + \beta b) = \gamma(a+b) = a'$ for some $a \varepsilon A$, $b \varepsilon B$. Since $ha + \beta b \varepsilon B'$ we have $\alpha a = a'$. Therefore α is onto. Let $\alpha : A \approx A'$ and let $b' \varepsilon B'$ be given. Then $\alpha a + (ha + \beta b) = b'$ for some $a \varepsilon A$, $b \varepsilon B$. Since $\alpha a \varepsilon A'$ we have $\alpha a = 0$. Therefore $a = 0$, $ha = 0$ and $\beta b = b'$. Therefore β is onto. Let $\beta : B \approx B'$ and let $\alpha a = 0$. Then $\gamma(a - \beta^{-1}ha) = \alpha a + (ha - ha) = 0$. Therefore $a - \beta^{-1}ha = 0$. Since $\beta^{-1}ha \varepsilon B$ it follows that $a = 0$. Therefore $\alpha : A \approx A'$. Thus α, β, γ are isomorphisms if (i) is satisfied.

Let α, β be isomorphisms. Then $\gamma = \gamma^* \delta$, where $\gamma^* : A + B \to A' + B'$, $\delta : A + B \to A + B$ are given by $\gamma^*(a+b) = \alpha a + \beta b$, $\delta(a+b) = a + (\beta^{-1}ha + b)$. Obviously γ^*, δ are isomorphisms and so therefore is γ. Moreover $g : M \to M$ is of the form (2.5), where

$$g(a+b+m) = \delta(a+b) + m \qquad (m \varepsilon M_{A+B}).$$

Let P be a permutation such that $PA' = A$, $PB' = B$. Let $f = f_\alpha$ be defined by (2.8) and let f_β, f_γ, f_{γ^*} be similarly defined in terms of β, γ; γ^* and the same permutation P. Then $f_\alpha b = b$, $f_\beta a = a$ and

$$f_\gamma(a+b) = P\alpha a + P(ha + \beta b) = f_\alpha + f_\beta(\beta^{-1}ha + b) = f_\alpha f_\beta g(a+b).$$

Since $g \varepsilon \Sigma$ it follows that

$$\tau(\gamma) = \tau(f_\gamma) = \tau(f_\alpha) + \tau(f_\beta) + \tau(g) = \tau(\alpha) + \tau(\beta)$$

and the proof is complete.

COROLLARY. *If any two of α, β, γ are simple isomorphisms, so is the third.*

Let $\theta : R \approx R$ be an automorphism of R and let $s_\theta : M \to M$ be the transformation which is given by

$$(2.9) \qquad s_\theta(r_1, r_2, \cdots) = (\theta r_1, \theta r_2, \cdots).$$

Obviously $s_{\theta^{-1}} = s_\theta^{-1}$ and $s_\theta s_\phi = s_{\theta\phi}$, where $\phi : R \approx R$. Also $s_\theta(rm) = (\theta r) s_\theta m$, where $r \varepsilon R$, $m \varepsilon M$. Hence it follows that, if $f : M \to M$ is an (operator) endomorphism, then $(s_\theta f s_\phi) rm = (\theta\phi r)(s_\theta f s_\phi) m$. Therefore $f^\theta r = r f^\theta$ where $f^\theta = s_\theta f s_\theta^{-1}$. Since $s_\theta m_i = m_i$ it follows that $f^\theta \varepsilon \mathfrak{A}$ if $f \varepsilon \mathfrak{A}$. Let $g : M \to M$ be given by (2.7). Since $s_\theta m = m_i$ we have $g^\theta m_i = (\theta\lambda) m_i + m_j$, $g^\theta m_k = m_k$. Therefore g^θ is also of the form (2.7) if $\theta\lambda \varepsilon \Lambda$. Clearly $(g_1 g_2)^\theta = g_1^\theta g_2^\theta$ and it follows that $f^\theta \varepsilon \Sigma$ if $f \varepsilon \Sigma$, provided $\theta\Lambda \subset \Lambda$.

We shall describe $\theta : R \approx R$ as a Λ-*automorphism* if, and only if, $\theta\Lambda = \Lambda$. The totality of Λ-automorphisms is obviously a group Θ. Since $f^\theta \varepsilon \Sigma$ if $f \varepsilon \Sigma$ and $\theta \varepsilon \Theta$ it follows that T admits Θ as a group of operators, according to the rule

$$(2.10) \qquad \theta\tau(f) = \tau(f^\theta).$$

Let $x \varepsilon R$ be any regular element, not necessarily an element of Λ, and let $\theta_x r = x r x^{-1}$. I say that

$$(2.11) \qquad \theta_x \tau = \tau$$

for each $\tau \varepsilon \mathrm{T}$. For let $f \varepsilon \mathfrak{A}$ be given by (2.1). Then

$$f^{\theta_x} m_i = s_{\theta_x} \Sigma_j f_{ij} m_j = \Sigma_j (x f_{ij} x^{-1}) m_j.$$

Let $f = (f)^n$ and let $g_x = (g_x)^n : M \to M$ be given by

$$g_x(r_1, \cdots, r_n, r_{n+1}, \cdots) = (r_1 x, \cdots, r_n x, r_{n+1}, \cdots).$$

Then $g_x m_i = x m_i$ if $i \leq n$. Since $fxm = xfm$,

$$f g_x m_i = x f m_i = \sum_{j=1}^{n} x f_{ij} m_j = \sum_{j=1}^{n} (x f_{ij} x^{-1}) x m_j = g_x f^{\theta_x} m_i \quad (i = 1, \cdots, n).$$

Therefore $f^{\theta_x} = g_x^{-1} f g_x$ and $\tau(f^{\theta_x}) = -\tau(g_x) + \tau(f) + \tau(g_x) = \tau(f)$, which proves (2.11).

Let $f \varepsilon \mathcal{A}$ and let $\mathbf{f}$ and $\mathbf{f}^n$ mean the same as in (2.2). Let g be given by (2.7) and let $\mathbf{g}$ be its matrix. Then $\mathbf{gf}$ is obtained from $\mathbf{f}$ by the following operations

(2.12)
a) *multiplying a row from the left by an element* $\lambda \varepsilon \Lambda$,
b) *adding a left multiple of one row to another, the multiplier being an arbitrary element* $r \varepsilon R$.

Therefore $f \varepsilon \Sigma$ if, and only if, $\mathbf{f} \rightarrow \mathbf{1}_\infty$ by a finite sequence of such transformations. Let $\mathbf{f} \rightarrow \mathbf{1}_\infty$ by such a sequence, $\sigma_1, \cdots, \sigma_p$, and let $\mathbf{f} = (f)^n$. Then there is a $k \geq 0$ such that no row of $\mathbf{f}$, after the $(n+k)$-th is involved in any of $\sigma_1, \cdots, \sigma_p$. Therefore $\sigma_1, \cdots, \sigma_p$ transform[13] $\mathbf{f}^{n+k}$ into $\mathbf{1}_{n+k}$, where

$$\mathbf{f}^{n+k} = \begin{bmatrix} \mathbf{f}^n & 0 \\ 0 & \mathbf{1}_k \end{bmatrix}.$$

Let R be the group ring of a group Γ and let Λ consist of the elements $\pm\gamma$, where $\gamma \varepsilon \Gamma$. If Γ is Abelian, the determinant, $|\mathbf{f}^n|$, of $\mathbf{f}^n$ can be calculated in the ordinary way. Obviously $|\mathbf{f}^n|$ is unaltered by (2.12b) or by an "expansion," $\mathbf{f}^n \rightarrow \mathbf{f}^{n+k}$, and a transformation of the form (2.12a) changes $|\mathbf{f}^n|$ into $\pm\gamma\,|\mathbf{f}^n|$. Therefore $\pm\,|\mathbf{f}^n|\,\varepsilon\,\Gamma$ if $f \varepsilon \Sigma$. Let Γ be cyclic of order 5 and let $\gamma \neq 1$. Then $(1-\gamma-\gamma^4)(1-\gamma^2-\gamma^3) = 1$. Therefore $f: M \rightarrow M$, given by $f(r_1, r_2, r_3, \cdots) = \{r_1(1-\gamma-\gamma^4), r_2, r_3, \cdots\}$ is in $\mathcal{A}$, but not in Σ. Therefore $\mathrm{T} \neq 0$. On the other hand it follows from the theory of integral, unimodular matrices, in case $\Gamma = 1$, and from Theorems 14, 15 in [11], that $\mathrm{T} = 0$ if Γ is of order 1, 2, 3, 4 or is cyclic infinite.

We continue, until Section **9**, without the assumption that R is a group ring.

3. Chain systems. By a *chain system*, $C = \{C_n\}$, we shall mean a family of basic modules, $C_n \subset M$, together with a boundary operator, $\partial = \{\partial_n\}$, which is a family of (operator) homomorphisms, $\partial_n : C_n \rightarrow C_{n-1}$, such that $\partial_n \partial_{n+1} = 0$. For the sake of completeness we define $\partial_0 C_0 = C_{-1} = 0$. Each C_n, being a basic module, is of finite rank. We do not require C_0 to be of rank 1, as we did in section **8** of CH(II). For example, we allow $C_0 = 0$. We assume that $C_n = 0$ for all sufficiently large values of n. If $C_n = 0$ when $n > N \geq 0$, but $C_N \neq 0$, we write $N = \dim C$. We write $C = 0$, and dim $C = -1$, if $C_n = 0$ for every $n \geq 0$. We insist that $C_p \cap C_q = 0$ if $p \neq q$ and C shall be the set-theoretic union of the groups $C_0, C_1, \cdots$. Thus $c \varepsilon C$ means that $c \varepsilon C_n$ and $c + c'$ is only defined if c, $c' \subset C_n$, for some $n \geq 0$. Also ∂ is a map, $\partial : C \rightarrow C$, of the set C into itself.

Until Section 9 we shall only consider chain mappings,[15] $f: C \to C'$, of C into a chain system, $C' = \{C'_n\}$, such that each $f: C_n \to C'_n$ is an operator homomorphism. That is to say, in the terminology of CH (II), f is associated with the identical isomorphism, $R \to R$. Thus[16] $\partial f = f\partial$, $fr = rf$, where $r \varepsilon R$ is an operator. Also $f \simeq g: C \to C'$ will mean that

$$g_n - f_n = \partial_{n+1}\eta_{n+1} + \eta_n\partial_n \qquad (n \geqq 0), \tag{3.1}$$

where $\eta = \{\eta_n\}$ is a chain deformation operator, and $f: C \equiv . C'$ will mean that there is a chain mapping, $f': C' \to C$, such that $f'f \simeq 1$, $ff' \simeq 1$ in the sense of (3.1). We shall call $f: C \to C'$ a simple isomorphism, and shall write $f: C \approx C'$ (Σ), if, and only if, it is a chain mapping such that $f_n: C_n \approx C'_n$ (Σ), for each $n \geq 0$.

Let B, C be given chain systems and let $B_n = B'_n + B''_n$, $C_n = C'_n + C''_n$, where, according to our convention, B'_n, B''_n, C'_n, C''_n are basic modules. Let $f: B \to C$ be a chain mapping such that

$$f_n(b' + b'') = f'_n b' + (g_n b' + f''_n b'') \qquad (b' \varepsilon B'_n, b'' \varepsilon B''_n),$$

for each $n \geq 0$, where $f'_n b' \varepsilon C'_n$, $g_n b', f''_n b'' \subset C''_n$.

Lemma 1. *If any two of* $\{f_n\}$, $\{f'_n\}$, $\{f''_n\}$ *are families of simple isomorphisms, so is the third.*

This follows immediately from the corollary to Theorem 2.

Let $C_n = C'_n + C''_n$ and let $\partial C'_n \subset C'_{n-1}$ for each $n \geqq 0$. Let $C' = \{C'_n\}$ and let $\partial': C' \to C'$ be defined by $\partial' c' = \partial c'$. Then $\partial'\partial' c' = \partial\partial c' = 0$. Under these, and only these conditions, we shall describe C', with the boundary operator ∂', as a *sub-system* of C. If also $\partial C''_n \subset C''_{n-1}$ $(n \geq 0)$, so that $\partial(c' + c'') = \partial' c' + \partial'' c''$ $(c' \varepsilon C'_n, c'' \varepsilon C''_n)$, then $C'' = \{C''_n\}$, with boundary operator ∂'', is also a sub-system. In this case we shall call C the *direct sum*, $C = C' + C'' = C'' + C'$, of C' and C''. Let C', C'' be given, disjoint,[17] chain systems. Then the direct sum, $C' + C''$, will be the system which consists of the groups $C'_n + C''_n$, with $\partial(c' + c'') = \partial' c' + \partial'' c''$. Similarly we define the direct sum of any finite set of disjoint chain systems.

---

[15] At this stage we do not impose any restriction such as $f_0 m_i = \lambda m_j$ on f_0, where m_i is a basis element of C_0. For example, $C \to 0$ is a chain mapping.

[16] We shall often use ∂ to denote the boundary operator in each of two or more systems, $C, C', C'', \cdots$, which occur in the same context. On other occasions we shall use $\partial, \partial', \partial'', \cdots$ to denote the boundary operators in $C, C, C'', \cdots$.

[17] We describe two or more basic modules, or chain systems, as disjoint if, and only if, $0 \varepsilon M$ is their only common element.

Let C' be a sub-system of C and let $C_n = C'_n + C''_n$. Let $j_n : C_n \to C''_n$ and $\partial''_n : C''_n \to C''_{n-1}$ be defined by

$$j_n(c' + c'') = c'', \qquad \partial''_n c'' = j_{n-1}\partial_n c''. \tag{3.2}$$

Then $\partial''_n j_n = j_{n-1}\partial_n$, since $\partial C' \subset C'$ and $j_{n-1}C'_{n-1} = 0$. Therefore

$$\partial''_n \partial''_{n+1} j_{n+1} = \partial_n j_n \partial_{n+1} = j_{n-1}\partial_n \partial_{n+1} = 0,$$

whence $\partial''_n \partial''_{n+1} = 0$. Therefore $C'' = \{C''_n\}$, with $\partial'' = \{\partial''_n\}$ as boundary operator, is a chain system. We call it the *residue system*, mod C′, and write $C'' = C - C'$. Notice, however, that an element in C'' is an element in the basic module C''_n, for some $n \geq 0$, not a residue class of elements in C. Notice also that $j = \{j_n\}$ is a chain mapping, $j : C \to C''$; also that $c - jc \varepsilon C'$, whence

$$\partial c'' - \partial'' c'' = \partial c'' - j\partial c'' \varepsilon C'. \tag{3.3}$$

Let B', C' be sub-systems of chain systems B, C and let $B'' = B - B'$, $C'' = C - C'$. Let $f : B \to C$ be a chain mapping such that $fB' \subset C'$. Then $f' : B' \to C'$, given by $f'b' = fb'$, is obviously a chain mapping. Let

$$f''_n = j_n f_n : B''_n \to C''_n.$$

Then $f''j = jf$, where $j : B \to B''$ is defined in the same way as $j : C \to C''$. Since $\partial'' j = j\partial$ we have

$$f''\partial'' j = jf\partial = j\partial f = \partial'' f'' j,$$

where ∂ operates on B, C and ∂'' on B'', C''. Therefore f'' is a chain mapping. We shall call $f' : B' \to C'$, $f'' : B'' \to C''$ the chain mappings *induced* by f. It follows from Lemma 1 that, if any two of f, f', f'' are simple isomorphisms, so is the third.

Let A be a common sub-system of B and C. Then we shall describe a chain mapping, $f : B \to C$, as *rel.* A if, and only if, $fa = a$ for each $a \varepsilon A$. We shall say that $f \simeq g : B \to C$, rel. A, if, and only if, $g - f = \partial\eta + \eta\partial$, where $\eta : B \to C$ is a deformation operator such that $\eta A = 0$.

Let $Z_n(C) = \partial_n^{-1}(0)$ and let

$$H_n(C) = Z_n(C) - \partial_{n+1}C_{n+1} \qquad (n \geq 0).$$

A chain mapping, $f : B \to C$, obviously induces a family of homomorphisms

$$f_* : H_n(B) \to H_n(C).$$

Let C' be a sub-system of C, let $i : C' \to C$ be the identical map, which is obviously a chain mapping, and let $j: C \rightarrow C''$ mean the same as before. Let $z'' \varepsilon Z_n(C'')$. Then it follows from (3.3) that $\partial z'' \varepsilon Z_{n-1}(C')$. Therefore ∂ induces a family of homomorphisms $d_*: H_n(C'') \rightarrow H_{n-1}(C')$, where $H_{-1}(C') = 0$. It is known[18] that the sequence of homomorphisms,

$$(3.4) \qquad \cdots \xrightarrow[d_*]{} H_n(C') \xrightarrow[i_*]{} H_n(C) \xrightarrow[j_*]{} H_n(C'') \xrightarrow[d_*]{} \cdots \xrightarrow[d_*]{} H_{-1}(C'),$$

is exact, meaning that the kernel of each homomorphism is the image group of its predecessor. We prove that $d_*H_n(C'') = i_*^{-1}(0)$. Let $\bar{z} \varepsilon H_n(X)$ ($X = C, C'$ or C'') be the residue class containing a given element $z \varepsilon Z_n(X)$. Let $z'' \varepsilon Z_n(C'')$. Then $\partial z'' = i\partial z'' \varepsilon Z_{n-1}(C)$, and

$$i_*d_*\bar{z}'' = i_*\overline{\partial z''} = \overline{i\partial z''} = \overline{\partial z''} = 0.$$

Therefore $d_*H_n(C'') \subset i_*^{-1}(0)$. Conversely, let $i_*\bar{z} = 0$, where $z' \varepsilon Z_{n-1}(C')$. This means that $iz' \varepsilon \partial C_n$, or that $z' = \partial c = \partial(c' + z'')$ $(c' \varepsilon C'_n, z'' \varepsilon C''_n)$. Therefore, writing $z' - \partial c' = z'_1$, we have

$$\bar{z}' = \bar{z}'_1 = \overline{\partial z''} = d_*\bar{z}'',$$

whence $d_*H_n(C') \doteq i_*^{-1}(0)$. It follows from similar arguments that $i_*H_n(C') = j_*^{-1}(0)$ and that $j_*H_n(C) = d_*^{-1}(0)$.

4. Deformation retracts. Let $C' \subset C$ be a sub-system and let $i: C \rightarrow C$ be the identical chain mapping. A chain mapping, $k: C \rightarrow C'$, will be called a *retraction* if, and only if, $ki = 1$. We shall call C' a *deformation retract* (D.R.) of C if, and only if, there is a retraction, $k: C \rightarrow C'$, such that $ik \simeq 1$, rel. C'. Let $ik \simeq 1$, rel. C', and let $k': C \rightarrow C'$ be any other retraction. Then $ik' \simeq ik'ik = ik \simeq 1$, rel. C'.

THEOREM 3. *A sub-system $C' \subset C$ is a D.R. of C if, and only if, $H_n(C - C') = 0$ for every $n \geqq 0$.*

Let C' be a D.R. of C and let $k: C \rightarrow C'$ be a retraction. Then

$$(4.1) \qquad 1 - ik = \partial\eta + \eta\partial,$$

where $\eta: C \rightarrow C$ is a deformation operator such that $\eta C' = 0$. Let $C'' = C - C'$ and let $z'' \varepsilon Z_n(C'')$. Then $\partial z'' \varepsilon C'$, whence $\eta\partial z'' = 0$. Clearly $ji = 0$, $jz'' = z''$, where $j: C \rightarrow C''$ is given by (3.2). Therefore

$$z'' = j(1 - ik)z'' = j(\partial\eta + \eta\partial)z'' = j\partial\eta z'' = \partial j\eta z''.$$

Therefore $H_n(C'') = 0$ $(n \geqq 0)$.

---

[18] See Theorem 3.3 in [15].

Conversely, let $H_n(C'') = 0$ for every $n \geqq 0$. Assume that there are homomorphisms,

$$k_r : C_r \to C'_r, \ \eta_{r+1} : C_r \to C_{r+1} \qquad (r = -1, \cdots, n)$$

such that $k_r i_r = 1$, $\partial_r k_r = k_{r-1}\partial_r$ and

$$(4.2)_r \qquad i_r k_r - 1 = \partial_{r+1}\eta_{r+1} + \eta_r \partial_r,$$

these conditions being vacuous if $n = 0$. Let $m''_1, \cdots, m''_p$ $(m''_\lambda = m_i \ )$ be the basis of C''_n and let

$$(4.3) \qquad (1 + \eta_n \partial_n) m''_\lambda = c'_\lambda + c''_\lambda \qquad (c'_\lambda \,\varepsilon\, C'_n, \ c''_\lambda = C''_n).$$

It follows from $(4.2)_{n-1}$ that

$$\partial_n(1 + \eta_n \partial_n) = (1 + \partial_n \eta_n)\partial_n = (i_{n-1}k_{n-1} - \eta_{n-1}\partial_{n-1})\partial_n = i_{n-1}k_{n-1}\partial_n.$$

Therefore $\partial_n(c'_\lambda + c''_\lambda) = i_{n-1}k_{n-1}\partial_n m''_\lambda = k_{n-1}\partial_n m''_\lambda$. Therefore $\partial''_n c''_\lambda = 0$. Since $H_n(C'') = 0$ we have $c''_\lambda = \partial''_{n+1} a''_\lambda$, for some $a''_\lambda \,\varepsilon\, C''_{n+1}$. Let $a'_\lambda = c''_\lambda - \partial_{n+1} a''_\lambda \,\varepsilon\, C'_n$. Then it follows from (4.3) that

$$(4.4) \qquad (1 + \eta_n \partial_n) m''_\lambda = c'_\lambda + a'_\lambda + \partial_{n+1} a''_\lambda.$$

Let $k_n : C_n \to C'_n$ and $\eta_{n+1} : C_n \to C_{n+1}$ be the operator homomorphisms defined by $k_n c' = c'$, $\eta_{n+1} c' = 0$ and

$$k_n m''_\lambda = c'_\lambda + a'_\lambda, \qquad \eta_{n+1} m''_\lambda = - a''_\lambda.$$

Then $(4.2)_n$ follows from (4.4). Also

$$\partial_n k_n c' = \partial_n c' = k_{n-1}\partial_n c'$$

$$\partial_n k_n m''_\lambda = \partial_n(c'_\lambda + a'_\lambda) = \partial_n(c'_\lambda + c''_\lambda) = k_{n-1}\partial_n m''_\lambda.$$

Therefore, starting with $k_{-1} = \eta_{-1} = 0$, the theorem follows by induction on n.

COROLLARY 1. *$C \equiv 0$ if, and only if, $H_n(C) = 0$ for every $n \geqq 0$.*

COROLLARY. 2. *C' is a D. R. of C if, and only if, $C - C' \equiv 0$.*

LEMMA 2. *If C' is a D. R. of C then $C \approx C' + C''$ (Σ), rel. C', where $C'' = C - C'$.*

Let $C^* = C' + C''$. Then $C^*_n = C_n$ and $\partial^* : C^* \to C^*$ is given by $\partial^*(c' + c'') = \partial' c' + \partial'' c''$. Let i, k, η mean the same as in (4.1) and let $f : C^* \to C$ be given by

$$f(c' + c'') = (c' - kc'') + c'' = c' + (\partial\eta + \eta\partial)c''.$$

Then $\partial fc' = \partial c' = f\partial^* c'$ and

$$\partial fc'' = \partial(\partial\eta + \eta\partial)c'' = \partial\eta\partial c'',$$
$$f\partial^* c'' = (\partial\eta + \eta\partial)\partial'' c'' = (\partial\eta + \eta\partial)(\partial c'' + c'_1) \qquad (c'_1 \varepsilon C').$$

Since $\partial C' \subset C'$ and $\eta C' = 0$ it follows that $f\partial^* c'' = \partial\eta\partial c'' = \partial fc''$. Therefore f is a chain mapping. It follows from Lemma 1 that $f: C^* \approx C\ (\Sigma)$ and the lemma is proved.

5. Simple equivalence. We shall describe a chain system, B, as *elementary* if, and only if, $B_n = 0$ when $n \neq r-1, r$, for some $r \geq 1$, and $\partial_r: B_r \approx B_{r-1}\ (\Sigma)$. This being so, it is obvious that $H_n(B) = 0$ for every $n \geq 0$. Therefore $B \equiv 0$, by Theorem 3, Corollary 1. We shall describe B as *collapsible* if, and only if, it is the direct sum of a finite set of elementary systems. Clearly $B \equiv 0$ if B is collapsible. It is obvious that B' is collapsible if $B' \approx B\ (\Sigma)$, where B is collapsible; also that, if B, B' are disjoint and collapsible, then $B + B'$ is collapsible; also that the direct sum of a set of r-dimensional elementary systems is itself elementary.

Let $B_n = A_n + Z_n$, let $\partial'_n: A_n \approx Z_{n-1}\ (\Sigma)$ and let $\partial_n: B_n \to B_{n-1}$ be given by $\partial_n a = \partial'_n a$, $\partial_n Z_n = 0$, $(n = 1, 2, \cdots)$. Then $B = \{B_n\}$, with $\partial = \{\partial_n\}$ as boundary operator, is the direct sum of the elementary systems $(\cdots, 0, A_n, Z_{n-1}, 0, \cdots)$. Therefore B is collapsible and any collapsible system is obviously of this form.

We shall say that C, C' are in the same *simple equivalence class*, and shall write $C \equiv C'\ (\Sigma)$ if, and only if, there are collapsible systems, B, B', such that

$$(5.1) \qquad f: B + C \approx B' + C'\ (\Sigma).$$

This being so, it follows from Theorem 3, Corollary 2, that C, C' are D. R.'s of $B + C$, $B' + C'$. Let

$$i: C \to B + C, \qquad i': C' \to B' + C'$$
$$k: B + C \to C, \qquad k': B' + C' \to C'$$

be the identity maps and any retractions. Let

$$(5.2) \qquad g \simeq k'fi: C \to C'$$

and let $g' \simeq kf^{-1}i': C' \to C$. Then

$$(5.3) \qquad g'g \simeq kf^{-1}i'k'fi \simeq kf^{-1}fi = 1.$$

Similarly $gg' \simeq 1$. Therefore $g: C \equiv C'$. We shall describe a chain mapping, $g: C \to C'$, as a *simple equivalence* and shall write $g: C \equiv C'$ (Σ), if, and only if, it is related by (5. 2) to some simple isomorphism of the type (5. 1). It follows from (5. 3) that, if $g: C \equiv C'$ (Σ) and if $g'': C' \to C$ is such that $gg'' \simeq 1$, then g'' is a simple equivalence. Obviously $g: C \equiv C'$ (Σ) if $g: C \approx C'$ (Σ).

The relation $C \equiv C'$ (Σ) is obviously reflexive and symmetric. We proceed to prove that it is transitive; also that, if $g: C \equiv C'$ (Σ) and $g': C' \equiv C''$ (Σ), then $g'g: C \equiv C''$ (Σ). Let C, C' be related by (5. 1), let $f^*: B^* + C' \approx B'' + C''$ (Σ), where B^*, B'' are collapsible, and assume that

$$B' \cap B^* = B^* \cap (B + C) = B' \cap (B'' + C'') = 0. \tag{5.4}$$

Then $B + B^* + C \approx B' + B^* + C' \approx B' + B'' + C''$ (Σ), whence $C \equiv C''$ (Σ). If (5. 4) are not satisfied we apply a permutation, $P_n: B'_n \to A'_n$, to each module B'_n, thus transforming B' into a new system, A', such that $P: B' \approx A'$ (Σ), where $P = \{P_n\}$, and $A' \cap C = 0$. Let

$$h: B' + C' \approx A' + C' \ (\Sigma)$$

be given by $h(b' + c') = Pb' + c'$. Then

$$hf: B + C \approx A' + C' \ (\Sigma). \tag{5.5}$$

Similarly we can replace B^* by $A^* = P^*B^*$. We can choose the basic modules A'_n, A^*_n in such a way that (5. 4) are satisfied when B', B^* are replaced by A', A^*. Therefore $C \equiv C''$ (Σ), and it follows that $C \equiv C'$ (Σ) is an equivalence relation.

Let $g: C \to C'$ be given by (5. 2) and let h mean the same as in (5. 5). Then $g \simeq (k'h^{-1})(hf)i: C \to C'$ and $k'h^{-1}: A' + C' \to C'$ is obviously a retraction. Also $k'h^{-1}$ can be extended to a retraction, $A' + A^* + C' \to C'$, by mapping A^* on zero. Therefore, if $g: C \equiv C'$ (Σ) and $g': C' \equiv C''$ (Σ) we lose no generality in assuming that g satisfies (5. 2) and that

$$f': B' + C' \approx B'' + C'' \ (\Sigma), \qquad g \simeq k''f'i': C' \to C'',$$

where $k'': B'' + C'' \to C''$ is a retraction. This being so,

$$f'f: B + C \approx B'' + C'' \ (\Sigma)$$

and

$$g'g \simeq k''f'i'k'fi \simeq k''f'fi: C \to C''.$$

Therefore $g'g: C \equiv C''$ (Σ).

A non-zero element, which is common to two chain systems, will be called an *accidental intersection,* unless it is in a common sub-system, which is explicitly mentioned in the context. Accidental intersections between any finite set of systems, $C, \cdots$, can always be eliminated, as in the paragraph containing (5.5), by replacing $C, \cdots$, by a set of chain systems, $PC, \cdots$, where $P_n : C_n \to (PC)_n$, is a suitable set of permutations. When the context requires it, we shall always assume that this has already been done.

Let $C \equiv C' \ (\Sigma)$, $C^* \equiv C'^* \ (\Sigma)$. Then $B + C \approx B' + C' \ (\Sigma)$, $B^* + C^* \approx B'^* + C'^* \ (\Sigma)$, where B, B' etc. are collapsible. Therefore, in the absence of accidental intersections, it follows from Lemma 1, in Section **3**, that

$$B + B^* + C + C^* \approx B' + B'^* + C' + C'^* \ (\Sigma),$$

whence

$$C + C^* \equiv C' + C'^* \ (\Sigma). \tag{5.6}$$

Let A be a common sub-system of C and C'. We shall write $C \equiv C'' \ (\Sigma)$, rel. A, if, and only if,

$$f : B + C \approx B' + C' \ (\Sigma), \text{rel. } A, \tag{5.7}$$

where B, B' are collapsible.

THEOREM 4. a) *If* $C \equiv C' \ (\Sigma)$, *rel.* A, *then* $C - A \equiv C' - A' \ (\Sigma)$.

b) *If* $C - A \equiv 0 \ (\Sigma)$, *then* $C \equiv A \ (\Sigma)$, *rel.* A.

Let C, C' be related by (5.7). Since f induces the identity, $A \to A \ (\Sigma)$, it follows from Lemma 1 that

$$f' : (B + C) - A \approx (B' + C') - A \ (\Sigma),$$

where f' is the chain mapping induced by f. Obviously

$$(B + C) - A = B + (C - A), \qquad (B' + C') - A = B' + (C' - A),$$

which proves (a).

Let $C'' \equiv 0 \ (\Sigma)$, where $C'' = C - A$. Then $B + C'' \equiv B' \ (\Sigma)$, where B, B' are collapsible. Since $C'' \equiv 0$ it follows from Theorem 3, Corollary 2, that A is a D. R. of C. Therefore $C \approx A + C'' \ (\Sigma)$, rel. A, by Lemma 2. Therefore

$$B + C \approx B + C'' + A \approx B' + A \ (\Sigma), \text{rel. } A.$$

Therefore $C \equiv A \ (\Sigma)$, rel. A, and the theorem is proved.

By a (p, q)-system, C, we shall mean a chain system such that $C_n = 0$ if $n < p$ or if $n > q$ $(p \leq q)$.

LEMMA[19] 3. *If* $C \equiv C' \ (\Sigma)$, *where* C, C' *are* (p,q)*-systems, there are collapsible* (p,q)*-systems,* B, B', *such that* $B + C \approx B' + C' \ (\Sigma)$.

Let $f: A + C \approx A' + C' \ (\Sigma)$, where A, A' are collapsible. Let $n = \dim A > q$. Then $n = \dim(A + C) = \dim(A' + C')$. It follows from the definition of a collapsible system that

$$(5.8) \qquad A = B^1 + \cdots + B^m,$$

where each B^i is an elementary system. Let E be the direct sum of all the n-dimensional summands, B^i, and let D be the direct sum of the others. Let $D', E' \subset A'$ be similarly defined. Then $E_n = (A + C)_n$, $E'_n = (A' + C')_n$, and $\partial_n : E_n \approx E_{n-1} \ (\Sigma)$, $\partial'_n : E'_n \approx E'_{n-1} \ (\Sigma)$ since E, E' are elementary systems. Also $f_n : E_n \approx E'_n \ (\Sigma)$. Therefore

$$fE_{n-1} = f\partial E_n = \partial' f E_n = \partial' E'_n = E'_{n-1}$$

and

$$f_{n-1} = \partial' f_n \partial_n^{-1} : E_{n-1} \approx E'_{n-1} \ (\Sigma).$$

Therefore $f: A + C \approx A' + C' \ (\Sigma)$ induces a simple isomorphism $E \approx E' \ (\Sigma)$. Since $A + C - E = D + C$, $A' + C' - E' = D' + C'$, it follows from Lemma 1 that

$$(5.9) \qquad D + C \approx D' + C' \ (\Sigma).$$

Now let $A_r \neq 0$ for some $r < p$ and let s be the least value of r with this property. Since $C'_s = 0$ and $A + C \approx A' + C' \ (\Sigma)$ it follows that s has the same property in A'. Let E now denote the direct sum of the $(s+1)$-dimensional summands in (5.8) and let D be the direct sum of the others. Let $D', E' \subset A'$ be similarly defined. Then $D_s = D'_s = 0$, by the minimal property of s. Therefore

$$(5.10) \qquad (D + C)_s = (D' + C')_s = 0.$$

Since E, E' are elementary systems we have

$$(5.11) \qquad \partial : E_{s+1} \approx E_s \ (\Sigma), \qquad \partial' : E'_{s+1} \approx E'_s \ (\Sigma).$$

Let $c \in C_{s+1}$, $d \in D_{s+1}$ and let $f(d + c) = e' + d' + c'$. Since $\partial'(d + c') = 0$, in consequence of (5.10), we have $\partial' e' = \partial'(e' + d' + c') = \partial' f(d + c) = f\partial(d + c) = 0$. Therefore it follows from (5.11) that $e' = 0$. Therefore $f(D + C) \subset D' + C'$. Let $h: E \to E'$ be the chain mapping induced by f. Then it follows from (5.10) that $h_s = f_s : E_s \approx E'_s \ (\Sigma)$ and also, since $he - fe \in (D' + C')_{s+1}$ if $e \in E_{s+1}$, that

---

[19] Cf. Theorem 1 on p. 1202 of [3].

$$\partial' h_{s+1} e = \partial' f_{s+1} e = f_s \partial e = h_s \partial e.$$

Therefore

$$h_{s+1} = \partial_{E'}^{-1} h_s \partial_E : E_{s+1} \approx E'_{s+1} \ (\Sigma),$$

where $\partial_E = \partial \mid E_{s+1}$, $\partial_{E'} = \partial' \mid E'_{s+1}$. Therefore $h : E \approx E' \ (\Sigma)$ and (5.9) again follows from Lemma 1. Lemma 3 now follows by induction on m in (5.8).

We are now approximately half way through the algebraic preliminaries. The simple homotopy equivalences will be defined as those which induce simple chain equivalences, in a sense explained in Section **10** below. But we have still to relate chain equivalences to the group T, which is defined in Section **2** above. The first step in this is to associate an element, $\tau(C) \varepsilon \mathrm{T}$, with each system, C, such that $C \equiv 0$. We shall do this by transforming C into an $(m, m+1)$-system, C^m, in which $\partial_{m+1} : C^m_{m+1} \approx C^m_m$, and defining $\tau(C) = (-1)^m \tau(\partial_{m+1})$. In Section **8** below we define a chain system called the "mapping cylinder," C^*, of a given chain equivalence $f : C \equiv C'$. This contains C as a sub-system and $C^* - C \equiv 0$. We define $\tau(f) = \tau(C^* - C)$. We shall also need to consider the effect of a Λ-automorphism, $\theta : R \approx R$, operating on T, because the chain mapping induced by a homotopy equivalence, $\phi : K \equiv K'$, is "associated" with an isomorphism $\pi_1(K) \approx \pi_1(K')$, namely the one induced by ϕ.

6. Null-equivalent systems. Let $C \equiv 0$. Then $k \simeq 1 : C \to C$, where $kC = 0$. Therefore there is a chain deformation operator, $\eta : C \to C$, such that

$$(6.1) \qquad \partial \eta + \eta \partial = 1.$$

Let $\delta = \eta \partial \eta$. That is to say, $\delta = \{\delta_n\}$, where $\delta_n = \eta_n \partial_n \eta_n : C_{n-1} \to C_n$. Therefore δ is also a chain deformation operator. It follows from (6.1) that $\partial \eta \partial = (1 - \eta \partial) \partial = \partial$. Therefore $\partial \delta + \delta \partial = \partial \eta \partial \eta + \eta \partial \eta \partial = \partial \eta + \eta \partial = 1$. Also $\partial \eta \eta = (1 - \eta \partial) \eta = \eta (1 - \partial \eta) = \eta \eta \partial$. Therefore $\delta \delta = \eta \partial \eta \eta \partial \eta = \eta \eta \eta \partial \partial \eta = 0$. Thus

$$(6.2) \qquad \partial \delta + \delta \partial = 1, \qquad \delta \delta = 0.$$

Let $P_1, P_2 : M \to M$ be permutations and let B be the elementary system in which

$$B_0 = 0, \quad B_i = P_i C_0, \quad B_n = 0 \qquad (i = 1, 2; n > 2)$$

and $\partial P_2 c_0 = P_1 c_0$ $(c_0 \varepsilon C_0)$. Let $C' = B + C$. Then $C'_n = C_n$ if $n \neq 1$ or 2 and

$$\left\{\begin{array}{ll} C'_1 = P_1C_0 + C_1, & C'_2 = P_2C_0 + C_0 \\ \partial'_1(P_1c_0 + c_1) = \partial_1 c_1, & \partial'_2(P_2c_0 + c_2) = P_1c_0 + \partial_2 c_2, \end{array}\right.$$

where $c_i \varepsilon C_i$. Let C^* be the system which consists of the same groups, $C^*_n = C'_n$, with $\partial^*_n = \partial_n$ if $n > 2$ and

$$(6.3) \qquad \partial^*_1(P_1c_0 + c_1) = c_0, \qquad \partial^*_2(P_2c_0 + c_2) = \delta_1 c_0 + \partial_2 c_2.$$

Let $f: C' \to C^*$ be given by $f_n = 1$ if $n \neq 1$ and $f_1(P_1c_0 + c_1) = P_1\partial_1 c_1 + (\delta_1 c_0 + c_1)$. Then

$$\begin{aligned} \partial^*_1 f_1(P_1c_0 + c_1) &= \partial_1 c_1 = f_0\partial'_1(P_1c_0 + c_1) \\ \partial^*_2 f_2(P_2c_0 + c_2) &= \delta_1 c_0 + \partial_2 c_2 = P_1\partial_1\partial_2 c_2 + \delta_1 c_0 + \partial_2 c_2 \\ &= f_1(P_1c_0 + \partial_2 c_2) = f_1\partial'_2(P_2c_0 + c_2) \end{aligned}$$

and $\partial^*_n f_n - f_{n-1}\partial'_n = \partial_n - \partial_n = 0$ if $n > 2$. Therefore f is a chain mapping. Let $g_1, h_1 : C'_1 \to C'_1$ be given by

$$\begin{aligned} g_1(P_1c_0 + c_1) &= P_1(c_0 + \partial_1 c_1) + c_1 \\ h_1(P_1c_0 + c_1) &= -P_1c_0 + (\delta_1 c_0 + c_1). \end{aligned}$$

Since $\partial_1\delta_1 = \partial_1\delta + \delta_0\partial_0 = 1$ we have

$$\begin{aligned} g_1h_1(P_1c_0 + c_1) &= g_1\{-P_1c_0 + (\delta_1c_0 + c_1)\} \\ &= P_1(-c_0 + c_0 + \partial_1 c_1) + (\delta_1 c_0 + c_1) = f_1(P_1c_0 + c_1). \end{aligned}$$

Therefore $f_1 = g_1h_1$. It follows from Theorem 2 in Section 2 that $g_1, h_1 : C'_1 \approx C'_1$ (Σ) and hence that $f: C \approx C'$ (Σ).

It follows from (6.3) that $C^* = B' + C^1$, where $B'_n = 0$ if $n > 1$,

$$B'_0 = C_0, B'_1 = P_1C_0, (\partial^* \mid B'_1) = P_1^{-1} : B'_1 \approx B'_0 \ (\Sigma)$$

and

$$(6.4) \qquad C^1_0 = 0, C^1_1 = C_1, C^1_2 = P_2C_0 + C_2, C^1_n = C_n \qquad (n > 2)$$

with $\partial^1 : C^1 \to C^1$ given by $\partial^1_n = \partial_n$ if $n > 2$ and

$$(6.5) \qquad \partial^1_2(P_2c_0 + c_2) = \delta_1 c_0 + \partial_2 c_2.$$

Let $m \geq 1$ and assume that there is a system, C^m, such that $C \equiv C^m$ (Σ) and

$$C^m_0 = \cdots = C^m_{m-1} = 0, \qquad C^m_{m+p+1} = C_{m+p+1} \qquad (p > 0).$$

Then it follows from the above argument, with C^m_{m+n} playing the part of C_n, that $C^m \equiv C^{m+1}$ (Σ), where C^{m+1} satisfies the same conditions as C^m, with m

replaced by $m+1$. It follows by induction on m that there is such a system for each value of m. Moreover $C^m \equiv C \equiv 0$. Therefore equations analogous to (6.2) are satisfied in C^m.

Let $N = \dim C$ and let $m \geqq N-1$. Then $C^m{}_n = 0$ unless $n = m$ or $m+1$. Therefore (6.2) reduces to

$$(6.6) \qquad \left\{ \begin{array}{l} \partial_{m+1}\delta_{m+1} = 1 : C^m{}_m \to C^m{}_m \\ \delta_{m+1}\partial_{m+1} = 1 : C^m{}_{m+1} \to C^m{}_{m+1}. \end{array} \right.$$

Therefore $\partial_{m+1} : C^m{}_{m+1} \approx C^m{}_m$ and $\delta_{m+1} = \partial^{-1}{}_{m+1}$. We define the *torsion*, $\tau(C)$, of C as

$$(6.7) \qquad \tau(C) = (-1)^m \tau(\partial_{m+1}) = (-1)^{m+1} \tau(\delta_{m+1}).$$

In (6.4) and (6.5) let C, C^1 be replaced by C^m, C^{m+1}, with $C^m{}_n = 0$ if $n > m+1$. Then (6.4), (6.5) become

$$C^{m+1}{}_{m+1} = C^m{}_{m+1}, \qquad C^m{}_{m+2} = PC^m{}_m,$$

where P is a permutation, and

$$\partial_{m+2} = \delta_{m+1} P^{-1} : C^{m+1}{}_{m+2} \to C^{m+1}{}_{m+1}.$$

Since $\tau(P) = 0$ it follows from (6.7) that $\tau(C) = (-1)^{m+1}\tau(\delta_{m+1}) = (-1)^{m+1}\tau(\partial_{m+2})$. Therefore $\tau(C)$ does not depend on the choice of $m \geqq N-1$. However $\tau(C)$ appears to depend on the particular choice of δ in (6.2) and on the construction for C^m. The following theorem shows that it does not.

THEOREM 5. *$\tau(C)$ depends only on C. Also $\tau(C) = \tau(C')$, if and only if, $C \equiv C'$ (Σ), given that $C \equiv C' \equiv 0$.*

Let $C \equiv C'$ (Σ), where $C \equiv 0$, and let C^m, C'^m be any given $(m, m+1)$-systems such that $C^m \equiv C$ (Σ), $C'^m \equiv C'$ (Σ). Let $\tau(C)$ be defined by (6.7), where C^m is now this given system, and let $\tau(C')$ be similarly defined in terms of C'^m. In particular we may have $C = C'$. Therefore, when we have proved that $\tau(C) = \tau(C')$, it will follow that $\tau(C)$ depends only on C and also that $\tau(C) = \tau(C')$ if $C \equiv C'$ (Σ).

By Lemma 3 $f : B + C^m \approx B' + C'^m$ (Σ), where B, B' are collapsible, and hence elementary $(m, m+1)$-systems. It follows from Theorem 3, Corollary 2, that $B + C^m \equiv B' + C'^m \equiv 0$. Therefore it follows from relations analogous to (6.6) that

$$\partial : (B + C^m)_{m+1} \approx (B + C^m)_m$$
$$\partial' : (B' + C'^m)_{m+1} \approx (B' + C'^m)_m.$$

P14711

Moreover $\partial = f_m^{-1}\partial' f_{m+1}$ since $f\partial = \partial' f$. Since f_m, f_{m+1} are simple isomorphisms it follows that $\tau(\partial) = \tau(\partial')$. Also

$$\begin{cases} \partial(b + c) = \partial_B b + \partial_{m+1} c & (b \in B_{m+1},\ c \in C^m_{m+1}) \\ \partial'(b' + c') = \partial_{B'} b' + \partial'_{m+1} c', \end{cases}$$

where $\partial_B = \partial \mid B_{m+1}$, $\partial_{B'} = \partial' \mid B'_{m+1}$. By the definition of an elementary system, $\tau(\partial_B) = \tau(\partial_{B'}) = 0$ and it follows from Theorem 2, in Section **2**, that $\tau(\partial_{m+1}) = \tau(\partial) = \tau(\partial') = \tau(\partial'_{m+1})$. Therefore $\tau(C) = \tau(C')$.

Conversely let $\tau(C) = \tau(C')$ and let $C^m \equiv C\ (\Sigma)$, $C'^m \equiv C'\ (\Sigma)$, where C^m, C'^m are $(m, m+1)$-systems. Let $C^m{}_m$ be of rank p and $C'^m{}_m$ of rank p'. If $p \neq p'$, say $p < p'$, we replace C^m by $B + C^m$, where B is an elementary $(m, m+1)$-system, such that B_m is of rank $p' - p$. Therefore we assume that $p = p'$. Moreover, after applying suitable permutations to $C^m{}_m$, $C^m{}_{m+1}$ we assume that $C^m{}_m = C'^m{}_m$, $C^m{}_{m+1} = C'^m{}_{m+1}$. Then $g = \partial'_{m+1}\partial^{-1}_{m+1} : C^m{}_m \approx C^m{}_m$ and $(-1)^m\tau(g) = \tau(C') - \tau(C) = 0$. Therefore $g : C^m{}_m \approx C^m{}_m\ (\Sigma)$. Let $f : C^m \to C'^m$ be given by $f_m = g$, $f_{m+1} = 1$. Then $\partial'_{m+1} f_{m+1} \equiv \partial'_{m+1} = f_m \partial_{m+1}$. Therefore $f : C^m \approx C'^m\ (\Sigma)$ and the theorem is proved.

Obviously $\tau(0) = 0$. Therefore we have the corollary:

COROLLARY. *$C \equiv 0\ (\Sigma)$ if, and only if, $\tau(C) = 0$.*

Let C' be a sub-system of C, and let $C'' = C - C'$.

THEOREM 6. *If any two of C, C', C'' are chain equivalent to 0, so is the third and $\tau(C) = \tau(C') + \tau(C'')$.*

Let X, Y, Z denote C, C', C'' in any order and let $X \equiv Y \equiv 0$. Then $H_n(X) = 0$, $H_n(Y) = 0$ for every $n \geq 0$, according to Theorem 3, Corollary 1. Therefore it follows from the exactness of the sequence (3.4) that $H_n(Z) = 0$, for every $n \geq 0$, and hence that $Z \equiv 0$.

Let $C \equiv C' \equiv C'' \equiv 0$. Then $C \approx C' + C''\ (\Sigma)$, according to Theorem 3, Corollary 2, and Lemma 2, and it follows from Theorem 5 that $\tau(C) = \tau(C' + C'')$. For a sufficiently large value of m we have $C' \equiv A'\ (\Sigma)$, $C'' = A''\ (\Sigma)$, where A', A'' are $(m, m+1)$-systems. Therefore it follows from (5.6) that $C' + C'' \equiv A' + A''\ (\Sigma)$ and from Theorem 5 that $\tau(C) = \tau(A' + A'')$. Let $\partial' : A'_{m+1} \to A_m$, $\partial'' : A''_{m+1} \to A''_m$, be the boundary homomorphisms, which are isomorphisms since $A' \equiv A'' \equiv 0$. Then it follows from Theorems 2 and 5 that

$$\tau(A' + A'') = (-1)^m\{\tau(\partial') + \tau(\partial'')\} = \tau(A') + \tau(A'') = \tau(C') + \tau(C'')$$

and the theorem is proved.

For purposes of calculation we exhibit the structure of the system, C^m, which is defined by reiterating the construction $C \to C^1$, leading to (6.4). To begin with we do not require $m+1 \geqq \dim C$. Let $m = 2k$ or $2k+1$ $(k \geqq 0)$ and let $D_0, D_1 \subset M$ be the basic modules

$$\left\{\begin{aligned} D_0 &= C_0 + C_1 + \cdots + C_r \\ D_1 &= C_1 + C_3 + \cdots + C_{2k+1}, \end{aligned}\right. \tag{6.8}$$

where $r = m$ if $m = 2k$, $r = m+1$ if $m = 2k+1$. Let $D = D_0 + D_1 = C_0 + C_1 + \cdots + C_{m+1}$ and let $\partial : D \to D$ be the homomorphism which is determined by $\partial : C \to C$. Let $\delta' : D \to D$ be the homomorphism which is given by

$$\begin{aligned} \delta' c &= \delta c \qquad \text{if } c \,\varepsilon\, C_s \qquad (s < m+1) \\ &= 0 \qquad \text{if } c \,\varepsilon\, C_{m+1}. \end{aligned}$$

Then $\delta'\delta' = 0$. Also it follows from (6.2) that

$$\begin{aligned} (\partial\delta' + \delta'\partial)c &= c \qquad \text{if } c \,\varepsilon\, C_s \qquad (s < m+1) \\ &= \delta\partial c \qquad \text{if } c \,\varepsilon\, C_{m+1}. \end{aligned} \tag{6.9}$$

Also $\partial D_i \subset D_j$, $\delta' D_i \subset D_j$ $(i \neq j;\, i, j = 0, 1)$. Let

$$\Delta = \partial + \delta' : (D, D_0, D_1) \to (D, D_1, D_0) \tag{6.10}$$

and let $\Delta_i : D_i \to D_j$ be the homomorphism which is induced by Δ. Then

$$\Delta_i \Delta_j d = \Delta_i \Delta d = \Delta\Delta d \qquad (d \,\varepsilon\, D_j). \tag{6.11}$$

Since $\partial\partial = \delta'\delta' = 0$ it follows from (6.9) that

$$\begin{aligned} \Delta\Delta c &= (\partial + \delta')(\partial + \delta')c \\ &= c \qquad \text{if } c \,\varepsilon\, C_s \ (s < m+1)\,; \ = \delta\partial c \qquad \text{if } c \,\varepsilon\, C_{m+1}. \end{aligned} \tag{6.12}$$

Also $\delta'\partial C_{m+2} \subset \delta' C_{m+1} = 0$, whence $\Delta\partial C_{m+2} = 0$.

Let $i = 0$, $j = 1$ if $m = 2k$ and let $i = 1$, $j = 0$ if $m = 2k+1$. Then $C_{m+1} \subset D_j$ and it follows from (6.11) and (6.12) that

$$\begin{aligned} &\Delta_j \Delta_i = 1 \\ &\Delta_i \Delta_j c = c \qquad \text{if } c \,\varepsilon\, C_s \ (s < m+1)\,; \ = \delta\partial c \qquad \text{if } c \,\varepsilon\, C_{m+1}. \end{aligned} \tag{6.13}$$

Let C^m, with boundary operator ∂^m, be the chain system, which is given by [20] $C^m{}_n = 0$ if $n < m$,

$$\left\{\begin{aligned} C^m{}_m &= D_i = \cdots + C_{m-2} + C_m \\ C^m{}_{m+1} &= D_j = \cdots + C_{m-1} + C_{m+1} \\ \partial^m{}_{m+1} &= \Delta_j, \end{aligned}\right. \tag{6.14}$$

---

[20] Cf. (5) on p. 205 of [10].

and $C^m{}_n = C_n$, $\partial^m{}_n = \partial_n$ if $n > m + 1$. Since $\Delta \partial C_{m+2} = 0$ it follows that $\partial^m \partial^m = 0$. Let δ^m be the deformation operator, which is given by

$$(6.15) \qquad \begin{aligned} &\delta^m{}_{m+1} = \Delta_i \\ &\delta^m{}_{m+2} c = 0 \qquad \text{if } c \,\varepsilon\, C_s \ (s < m+1)\,; \ = \delta_{m+2} c \qquad \text{if } c \,\varepsilon\, C_{m+1} \end{aligned}$$

and $\delta^m{}_n = \delta_n$ if $n > m + 2$. Then $\delta^m{}_{n+1} \delta^m{}_n = 0$ if $n > m + 1$ and

$$\delta^m{}_{m+2} \delta^m{}_{m+1} (\cdots + c_m) = \delta_{m+2} \delta_{m+1} c_m = 0,$$

where $c_m \,\varepsilon\, C_m$. Therefore $\delta^m \delta^m = 0$. Also it follows from (6.13) that

$$\partial^m{}_{m+1} \delta^m{}_{m+1} = 1, \qquad \partial^m{}_{m+2} \delta^m{}_{m+2} + \delta^m{}_{m+1} \partial^m{}_{m+1} = 1,$$

whence $\partial^m \delta^m + \delta^m \partial^m = 1$.

Let C^{m+1} be the system, with boundary operator ∂^{m+1}, which is obtained from C^m by the construction, $C \to C^1$, leading to (6.4), with $C^m{}_{m+n}$ playing the part of C_n and with $P_2 C_0$ replaced by C_0 and P_2 by 1 in (6.4), (6.5). Then

$$\begin{aligned} C^{m+1}{}_{m+1} &= C^m{}_{m+1} = \cdots + C_{m-1} + C_{m+1} \\ C^{m+1}{}_{m+2} &= C^m{}_m + C^m{}_{m+2} = \cdots + C_m + C_{m+2} \end{aligned}$$

and (6.5) becomes

$$\partial^{m+1}{}_{m+2}(c^m + c_{m+2}) = \partial^m{}_{m+1} c^m + \partial_{m+2} c_{m+2} = \Delta_i c^m + \partial_{m+2} c_{m+2} = \Delta^* (c^m + c_{m+2}),$$

where $c^m \,\varepsilon\, C^m{}_m$, $c_{m+2} \,\varepsilon\, C_{m+2}$ and Δ^* is defined by (6.10), with m replaced by $m + 1$. Therefore C^{m+1} is defined in the same way as C^m and we define δ^{m+1} by (6.15), with m replaced by $m + 1$. Starting with $C^0 = C$, it follows by induction on m that the construction $C \to C^1$, reiterated m times, leads to C^m. Therefore $C \equiv C^m$ (Σ). We now take $m \geq \dim C - 1$, thus giving an explicit definition of ∂_{m+1}, δ_{m+1} in (6.7).

Let $R' \subset R$ be a sub-ring, which is the image of R in a homomorphism, $\phi : R \to R'$, such that $\phi r' = r'$ if $r' \,\varepsilon\, R'$. Let $1 \,\varepsilon\, R'$ and let $\Lambda' = \phi \Lambda$. It is easily verified that Λ' is a (multiplicative) group . Let $M' \subset M$ be the sub-module, which consists of all the elements $(r'_1, r'_2, \cdots)$, where $r'_i \,\varepsilon\, R'$, and let R' be such that every admissible automorphism, $M' \to M'$, is Λ-simple. That is to say, every matrix of the form (2.2), with elements in R', can be transformed into the unit matrix by a finite sequence of the transformations (2.12), with $\lambda \,\varepsilon\, \Lambda$, $r \,\varepsilon\, R$.

Let $\psi : M \to M'$ be given by $\psi(r_1, r_2, \cdots) = (\phi r_1, \phi r_2, \cdots)$. Then $\psi m_i = m_i$ and $\psi(rm) = (\phi r) \psi m$. Let $f : M \approx M$ be an admissible automorphism, which is given by $f m_i = \Sigma_j f_{ij} m_j$, $(f_{ij} \,\varepsilon\, R)$. Then $\psi f m_i = \Sigma_j (\phi f_{ij}) m_j$ and it follows that $f_{ij} \,\varepsilon\, R'$ if (and only if) $\psi f = f \psi$. Therefore $f \,\varepsilon\, \Sigma$ if $\psi f = f \psi$.

Let C be a chain system, with boundary operator ∂, and let $\psi : M \to M'$ mean the same as before. Then $\psi C_r \subset C_r$, since $\psi m_i = m_i$, and $\partial\psi$, $\psi\partial$ are two families of homomorphisms $\partial\psi, \psi\partial : C_n \to C_{n-1}$.

Lemma 4. *If* $C \equiv 0$ *and if* $\partial\psi = \psi\partial$, *then* $C \equiv 0$ (Σ).

Let $C \equiv 0$ and let $\partial\psi = \psi\partial$. Let $\eta : C \to C$ be a deformation operator, which satisfies (6.1), and let $\xi : C \to C$ be the deformation operator determined by $\xi m_i = \psi\eta m_i$, $\xi r = r\xi$. Since $\partial\psi = \psi\partial$ and $\psi m_i = m_i$ we have $(\partial\xi + \xi\partial) m_i = \psi(\partial\eta + \eta\partial) m_i = m_i$. Therefore $\partial\xi + \xi\partial = 1$. Also $\psi\xi = \xi$, since $\psi\psi = \psi$. Therefore $\xi\psi m_i = \xi m_i = \psi\xi m_i$, whence $\xi\psi = \psi\xi$. Moreover $\delta = \xi\partial\xi$ is a deformation operator such that $\delta\psi = \psi\delta$, which satisfies (6.2). It follows from (6.5) and induction on m, or from the explicit formulae (6.14), that ∂_{m+1}, in (6.6), may be constructed so as to commute with ψ. Therefore ∂_{m+1} is a simple isomorphism. Therefore $\tau(C) = 0$ and the lemma follows from the corollary to Theorem 5.

Let R be the group ring of a group Γ, let Λ consist of the elements $\pm\gamma$ $(\gamma \varepsilon \Gamma)$ and let R' consist of the integral multiples of $1 \varepsilon \Gamma$. Let $\phi : R \to R'$ be given by $\phi\Gamma = 1$. Then we have the corollary:

Corollary. *Let* $C \equiv 0$, *let* $(m^n{}_1, \cdots, m^n{}_{p_n})$ *be the basis of* C_n $(m^n{}_i = m_{j_i})$ *and let*

$$\partial m^n{}_i = \Sigma_j d^n{}_{ij} m^{n-1}{}_j \qquad (n = 1, 2, \cdots),$$

where $d^n{}_{ij}$ *are integers. Then* $C \equiv 0$ (Σ).

7. Conjugate systems. Let C be a given chain system with boundary operator ∂. Let $\theta : R \approx R$ be a given automorphism and let $s_\theta : M \to M$ mean the same as in (2.9). We shall also use s_θ to denote the semi-linear transformation $s' : C_r \to C_r$, which is given by[21] $s'c = s_\theta c$ for each $c \varepsilon C_r$ $(r = 0, 1, \cdots)$. Let

$$\partial^\theta = s_\theta \partial s_\theta^{-1} : C_n \to C_{n-1} \qquad (n = 1, 2, \cdots).$$

Then $\partial^\theta\partial^\theta = s_\theta\partial\partial s_\theta^{-1} = 0$. Obviously $\partial^\theta r = r\partial^\theta$ for each operator $r \varepsilon R$. Therefore ∂^θ is a boundary operator. Let C^θ be the chain system, which consists of the modules C_n with the boundary operator ∂^θ. We shall describe C^θ as *conjugate* to C.

Let $f : C \to C'$ be a chain mapping, let $f^\theta{}_n = s_\theta f_n s_\theta^{-1} : C_n \to C'_n$, and let $f^\theta = \{f^\theta{}_n\}$. Obviously $f^\theta r = r f^\theta$ and $f^\theta\partial^\theta = s_\theta f \partial s_\theta^{-1} = s_\theta \partial f s_\theta^{-1} = \partial^\theta f^\theta$. Therefore $f^\theta : C^\theta \to C'^\theta$ is a chain mapping. On transforming the relevant equations by s_θ we see that, if $f: C \equiv C'$, then

---

[21] $s_\theta C_r = C_r$ since $s_\theta m_i = m_i$ and C_r is a basic module in M.

(7.1) $$f^\theta : C^\theta \equiv C'^\theta.$$

Let $C \equiv 0$ and let ∂, δ satisfy (6.2). Then ∂^θ and $\delta^\theta = s_\theta \delta s_\theta^{-1}$ obviously satisfy (6.2). It follows from (6.5) and induction on m, or from the explicit formulae (6.14), that the construction for C^m, with ∂, δ replaced by $\partial^\theta, \delta^\theta$, leads to the conjugate system $(C^m)^\theta$. Therefore

(7.2) $$\tau(C^\theta) = (-1)^m \tau(\partial^\theta_{m+1}).$$

Let $\theta\Lambda = \Lambda$. Then it follows from (7.2) that

(7.3) $$\tau(C^\theta) = \theta\tau(C),$$

where $\theta : \mathrm{T} \to \mathrm{T}$ is defined by (2.10).

8. Mapping cylinders. Let C, C' be disjoint chain systems, with boundary operators ∂, ∂', and let $f: C \to C'$ be a chain mapping. Let αC_{n-1} be the image of C_{n-1} in a simple isomorphism $\alpha : C_{n-} \approx \alpha C_{n-1}$, which is induced by a permutation $P_{n-1} : M \to M$. Then αC_{n-1} is a basic module. Let C^*_n be the direct sum $C^*_n = C'_n + C_n + \alpha C_{n-1}$. Let $\partial : C^*_n \to C^*_{n-1}$ be defined by

(8.1) $$\begin{cases} \text{a)} \quad \partial^* c = \partial c, \quad \partial^* c' = \partial' c' & (c \,\varepsilon\, C_n,\ c' \,\varepsilon\, C'_n) \\ \text{b)} \quad \partial^* \alpha c = (f - 1 - \alpha\partial) c & (c \,\varepsilon\, C_{n-1}). \end{cases}$$

We shall write (8.1b) as $\partial^*\alpha = f - 1 - \alpha\partial$, using 1, f as abbreviations for i, $i'f$, where $i : C_n \to C^*_n$, $i' : C'_n \to C^*_n$ are the identities. Obviously $\partial^*\partial^*(C'_n + C_n) = 0$ and $\partial^*\partial^*\alpha = \partial' f - \partial - \partial^*\alpha\partial = (f-1)\partial - (f - 1 - \alpha\partial)\partial = 0$. Therefore $C^* = \{C^*_n\}$, with ∂^* as boundary operator, is a chain system. We shall call it the *mapping cylinder* of f. Clearly C, C' are subsystems of C.

Lemma 5. *$C^* - C'$ is collapsible.*

Let $C'' = C^* - C'$. Then $C''_n = C_n + \alpha C_{n-1}$ and

(8.2) $$\partial''(c_2 + \alpha c_1) = \partial c_2 - (1 + \alpha\partial) c_1 = (\partial c_2 - c_1) - \alpha\partial c_1,$$

where $c_2 \,\varepsilon\, C_n$, $c_1 \,\varepsilon\, C_{n-1}$. Let $\partial^0 : C'' \to C''$ be given by

$$\partial^0 c = 0, \qquad \partial^0 \alpha c = c.$$

Then $\partial^0\partial^0 = 0$ and it follows that $\{C''_n\}$, with ∂^0 as boundary operator, is a chain system C^0, which is obviously collapsible. Let $g : C'' \to C^0$ be given by

$g(c_2 + \alpha c_1) = c_2 + \alpha(\partial c_2 - c_1)$. Then $g\partial''(c_2 + \alpha c_1) = g(\partial c_2 - c_1 - \alpha\partial c_1) = \partial c_2 - c_1 + \alpha(-\partial c_1 + \partial c_1) = \partial^0 g(c_2 + \alpha c_1)$. Therefore it follows from Lemma 1, Section **3**, that $g : C'' \approx C^0$ (Σ), and Lemma 5 is proved.

It follows from Lemma 5 and Theorem 4 that $C^* \equiv C'$ (Σ), rel. C'. Therefore C' is a D. R. of C^* and any retraction $k' : C^* \to C'$ is a simple equivalence. Let k' be given by $k'c = fc$, $k'c' = c'$, $k'\alpha c = 0$ $(c \in C, c' \in C')$. Then it follows from (8. 1) that $k'\partial^* = \partial' k'$. Therefore k' is a chain mapping, which is a retraction, since $k'c' = c'$. Also $k'c = fc = fic$. Therefore we have the corollary.

COROLLARY. *$k' : C^* \equiv C'$ (Σ), rel. C', and $f = k'i$.*

Let $C(f) = C^* - C$. Then $C_n(f) = C'_n + \alpha C_{n-1}$, and it follows from (8. 1) that the boundary operator, $\partial^f : C(f) \to C(f)$, is given by

$$\partial^f c' = \partial' c', \qquad \partial^f \alpha = f - \alpha\partial. \tag{8. 3}$$

LEMMA 6. *$C(f) \equiv 0$ if, and only if, $f : C \equiv C'$.*

Let $C(f) \equiv 0$. Then C is a D. R. of C^*, according to Theorem 3, Corollary 2. Therefore $i : C \equiv C^*$ and, by the corollary to Lemma 5, $k' : C^* \equiv C'$. Therefore $f : C \equiv C'$.

Conversely, let $f : C \equiv C'$. Then

$$f_* = k'_* i_* : H_n(C) \approx H_n(C') \qquad (n = 0, 1, \cdots)$$

and it follows that $i_* : H_n(C) \approx H_n(C^*)$. Therefore it follows from the exactness of the sequence

$$H_n(C) \to H_n(C^*) \to H_n\{C(f)\} \to H_{n-1}(C) \to H_{n-1}(C^*)$$

that $H_n\{C(f)\} = 0$ for every[22] $n \geq 0$. Therefore $C(f) \equiv 0$ by Theorem 3, Corollary 1.

LEMMA 7. *If $f \simeq g : C \to C'$, then $C(f) \approx C(g)$ (Σ).*

Let $g - f = \partial\eta + \eta\partial$, where $\eta : C \to C'$ is a deformation operator, and let βC_n and $\beta : C_n \approx \beta C_n$ be the analogues, in $C(g)$, of αC_n and α. Let $h : C(f) \to C(g)$ be given by $h(c' + \alpha c) = (c' - \eta c) + \beta c$, which we write as $hc' = c'$, $h\alpha = \beta - \eta$. Then $\partial h = h\partial$ in C' and since $hf = f$ we have[23]

$$\partial h\alpha = \partial(\beta - \eta) = g - \beta\partial - (g - f - \eta\partial) = f - (\beta - \eta)\partial = hf - h\alpha\partial = h\partial\alpha.$$

Therefore it follows from Lemma 1 that $h : C(f) \approx C(g)$ (Σ) and the lemma is proved.

---

[22] Cf. Section **3** in [5].

[23] We now use ∂ to denote the boundary operator in all our systems.

Let $f: C \equiv C'$. Then it follows from Lemma 6 that $C(f) \equiv 0$. We define $\tau(f) = \tau\{C(f)\}$ and call $\tau(f)$ the *torsion* of f. It follows from Lemma 7 that $\tau(f)$ depends only on the chain-homotopy class containing f and we shall also call it the *torsion* of this class.

Let $f: C \rightarrow C'$, $f': C' \rightarrow C''$ be any chain mappings.

LEMMA 8. *There is a chain system D, containing $C(f)$ as a sub-system, such that*[24] $D - C(f) = C(f')$ *and* $D \equiv C(f'f)$ (Σ).

Let $\alpha' C'_n \subset C(f')$ and $\alpha'' C_n \subset C(f'f)$ be the analogues of αC_n. Let C'^* be the mapping cylinder of f' and let D be the direct sum $D = C'^* + C(f)$ with the "united sub-system" C'. That is to say $D_n = C''_n + C'_n + \alpha' C'_{n-1} + \alpha C_{n-1}$ and $\partial: D \rightarrow D$ is determined by $\partial: C'^* \rightarrow C'^*$ and $\partial: C(f) \rightarrow C(f)$, which coincide in C'. Thus $\partial\alpha' = f' - 1 - \alpha'\partial$, $\partial\alpha = f - \alpha\partial$. Moreover $C(f)$ is a sub-system of D and obviously $D - C(f) = C(f')$.

Let D' be the direct sum $D' = C'^* + C(f'f)$, with the united sub-system C''. Then $D'_n = C''_n + C'_n + \alpha' C'_{n-1} + \alpha'' C_{n-1}$ and $\partial\alpha'' = f'f - \alpha''\partial$.

Let $g: D \rightarrow D'$ be given by

$$g(c'^* + \alpha c) = (c'^* - \alpha' f c) + \alpha'' c \qquad (c \in C, c'^* \in C'^*),$$

which we write as $gc'^* = c'^*$, $g\alpha = \alpha'' - \alpha' f$. Then $\partial g = g\partial$ in C'^*. Since $gfc = fc$ we have

$$\partial g\alpha = \partial\alpha'' - \partial\alpha' f = f'f - \alpha''\partial - (f' - 1 - \alpha'\partial)f = f - \alpha''\partial + \alpha'\partial f$$
$$= f - (\alpha'' - \alpha' f)\partial = g(f - \alpha\partial) = g\partial\alpha.$$

Therefore $g: D \approx D'$ (Σ). Clearly $D' - C(f'f) = C'^* - C''$ and it follows from Lemma 5 and Theorem 4(b) that $C(f'f) \equiv D' \approx D$ (Σ). This completes the proof.

Let $f: C \equiv C'$, $f': C' \equiv C''$.

THEOREM 7. $\tau(f'f) = \tau(f') + \tau(f)$.

Let D mean the same as in Lemma 8. Then it follows from Lemma 8 and Theorems 5, 6 that $\tau\{C(f'f)\} = \tau(D) = \tau\{C(f')\} + \tau\{C(f)\}$, and the theorem is proved.

LEMMA 9. *If $f: C \approx C'$ (Σ) then $\tau(f) = 0$.*

Let $f: C \approx C'$ (Σ) and let C^* be the mapping cylinder of f. Let $g: C(f) \rightarrow C^* - C'$ be given by $g(c' + \alpha c) = f^{-1}c' - \alpha c$. Then $\partial g c' = g\partial c'$ and it follows from (8.2) that

---

[24] I. e. $D - C(f) = C(f')$ if the permutations $\alpha': C'_n \rightarrow \alpha' C'_n \subset C(f')$ are suitably chosen.

$$\partial g\alpha = 1 + \alpha\partial = f^{-1}f + \alpha\partial = g(f - \alpha\partial) = g\partial\alpha.$$

Therefore $g: C(f) \approx C^{*} - C' \ (\Sigma)$ and the lemma follows from Lemma 5 and the corollary to Theorem 5.

Let $f: C \equiv C'$ and let us discard the (implicit) condition that $C \cap C' = 0$. Let $h: A \approx C \ (\Sigma)$, where A is a chain system such that $A \cap C' = 0$. Then $fh: A \equiv C'$ and we define $\tau(f)$ by $\tau(f) = \tau(fh)$. Let (h', A) be any other pair such that $h': A \approx C \ (\Sigma)$, $A' \cap C' = 0$. Let $h'': A'' \approx C \ (\Sigma)$, where A'' is disjoint from A, A', C'. Then $fh: A \equiv C'$, $h^{-1}h'': A'' \approx A \ (\Sigma)$, $fh'': A'' \equiv C'$. Therefore it follows from Theorem 7 and Lemma 9 that $\tau(fh'') = \tau(fh) + \tau(h^{-1}h'') = \tau(fh)$. Similarly $\tau(fh'') = \tau(fh')$. Therefore $\tau(f)$ is independent of the choice of h, A.

Let $f \simeq g: C \equiv C'$, where $C \cap C' \neq 0$. Then $fh \cong gh: A \equiv C'$, and in consequence of Lemma 7 we have:

THEOREM 8. *If* $f \simeq g: C \equiv C'$, *then* $\tau(f) = \tau(g)$.

Let $f: C \equiv C'$, $f': C' \equiv C''$. Let $h: A \approx C \ (\Sigma)$, $h': A' \approx C' \ (\Sigma)$, where A. A' are disjoint from C', C'' and from each other. Then $f'h': A' \equiv C''$, $h'^{-1}fh: A \equiv A'$, $f'fh: A \equiv C''$, and $h'^{-1}: C' \approx A' \ (\Sigma)$, $fh: A \equiv C'$. Therefore it follows from Theorem 7, and Lemma 9 that

$$\begin{aligned}\tau(f'f) &= \tau(f'fh) = \tau(f'h' \cdot h'^{-1}fh) = \tau(f'h') + \tau(h'^{-1}fh) \\ &= \tau(f'h') - \tau(h') + \tau(fh) = \tau(f') + \tau(f).\end{aligned}$$

Therefore Theorem 7 is valid, even when C, C', C'' are not disjoint from each other.

Similarly Lemma 9 is valid, even if $C \cap C' \neq 0$.

THEOREM 9. *Given* $g: C \equiv C'$, *then* $\tau(g) = 0$ *if, and only if,* $g: C \equiv C' \ (\Sigma)$.

It follows from Theorem 7 and Lemma 9 that we may assume $C \cap C' = 0$. This being so, let $\tau(g) = 0$ and let C^{*} be the mapping cylinder of g. Since $\tau\{C(g)\} = \tau(g) = 0$ it follows from Theorem 5 that $C(g) \equiv 0 \ (\Sigma)$. Therefore it follows from Theorem 4(b) that $C^{*} \equiv C \ (\Sigma)$, rel. C, whence $i: C \equiv C^{*} \ (\Sigma)$. Therefore it follows from Corollary to Lemma 5 that $g: C \equiv C' \ (\Sigma)$.

Conversely, let $g: C \equiv C' \ (\Sigma)$. Then $f: B + C \approx B' + C' \ (\Sigma)$, $g \simeq k'fi$, where B, B' are collapsible systems, $i: C \to B + C$ is the identity and $k': B' + C' \to C'$ is a retraction. Assume that $\tau(i) = \tau(i') = 0$, where $i': C' \to B' + C'$ is the identity. Then $\tau(k') = 0$, since $\tau(k') + \tau(i') = \tau(k'i') = \tau(1) = 0$. Also $\tau(f) = 0$, according to Lemma 9, and it follows from Theorems 7, 8 that $\tau(g) = 0$.

CUTTA

It remains to prove that $\tau(i) = \tau(i') = 0$. Let $h: A \approx C\ (\Sigma)$, where $A \cap (B + C) = 0$. Since $ihA = C$ it follows from (8.3) that $C(ih) = B + C(h)$, and from Theorems 5, 6 and Lemma 6 that $\tau\{C(ih)\} = \tau(B) + \tau\{C(h)\} = 0$. Therefore $\tau(i) = \tau(ih) = 0$ and similarly $\tau(i') = 0$. This completes the proof.

Let A, A' be sub-systems of C, C' and let $h: C \rightarrow C'$ be a chain mapping such that $hA \subset A'$. Let $B = C - A$, $B' = C' - A'$ and let $f: A \rightarrow A'$, $g: B \rightarrow B'$, be the chain mappings induced by h.

THEOREM 10. *If any two of f, g, h are chain equivalences so is the third, and* $\tau(h) = \tau(f) + \tau(g)$.

Assuming that $C \cap C' = 0$ we have

$$C_n(h) = C'_n + \alpha C_{n-1} = A'_n + B'_n + \alpha A_{n-1} + \alpha B_{n-1} = C_n(f) + C_n(g).$$

Let $D = C(h) - C(f)$. Then $D_n = C_n(g)$ and I say that $D = C(g)$. For let ∂_X denote the boundary operator in X, where X stands for any of the systems $C, C', \cdots$. Since $A' \subset C(f)$ we have[25]

$$\partial_D b' \equiv \partial_{C(h)} b' = \partial_{C'} b' \equiv \partial_{B'} b' = \partial_{C(g)} b', \qquad \text{mod. } C(f),$$

where $b' \varepsilon B'$. Since $\alpha A \subset C(f)$ we have $\alpha c_1 \equiv \alpha c_2$, mod. $C(f)$, if $c_1 \equiv c_2$, mod. A, where $c_1, c_2 \subset C$. Therefore

$$\partial_D \alpha b \equiv \partial_{C(h)} \alpha b = hb - \alpha \partial_C b \equiv gb - \alpha \partial_B b = \partial_{C(g)} b, \qquad \text{mod. } C(f),$$

where $b \varepsilon$ B. Therefore $\partial_D d \equiv \partial_{C(g)} d$, mod. $C(f)$, for any $d \varepsilon D$. But $D \cap C(f) = 0$. Therefore $\partial_D = \partial_{C(g)}$, whence $D = C(g)$. The theorem now follows from Lemma 6 and Theorem 6.

In consequence of Theorem 6 we have:

COROLLARY. *If any two of f, g, h are simple equivalences, so is the third.*

Using the same notation as in Theorem 10, let $A' = A = C \cap C'$ and let $h: C \rightarrow C'$ be rel. A. Then we define the *mapping cylinder*, C^*, of h in the same way as when $A = 0$, except that $C^*_n = C'_n + B_n + \alpha B_{n-1} = B'_n + C_n + \alpha B_{n-1}$ where $\alpha: B_n \approx \alpha B_n\ (\Sigma)$ is induced by a permutation, $P_n: M \rightarrow M$, and ∂^* is given by (8.1), with $c \varepsilon B$ and $\partial = \partial_B$ in (8.1b). Obviously $C^* - C = C(g)$. Therefore it follows from Theorem 10, with $f = 1$, and from Lemma 6, that $h: C \equiv C'$ if, and only if, $C^* - C \equiv 0$. It follows from the corollary to Theorem 10 and Theorem 9 that h is a simple equivalence if, and only if,

---

[25] If Y is a sub-system of X, then $x \equiv x'$, mod $Y (x, x' \subset X)$ means that $x - x' \varepsilon Y$. Clearly $\partial_z x \equiv \partial_x x$, mod Y, where $Z = X - Y$.

$$C^* - C \equiv 0 \ (\Sigma). \tag{8.4}$$

Let $f : C \equiv C'$ and let $\theta : R \approx R$ be any Λ-automorphism. Then $f^\theta : C^\theta \equiv C'^\theta$, according to (7.1). Since α, in (8.1), is the isomorphism induced by a permutation it follows that $\alpha s_\theta = s_\theta \alpha$. Therefore it follows from (8.3) that $C(f^\theta) = C(f)^\theta$ and from (7.3) that

$$\tau(f^\theta) = \theta\tau(f). \tag{8.5}$$

Let $f : C \equiv C'$. In order to calculate $\tau(f)$ we need to know a deformation operator, $\xi : C(f) \rightarrow C(f)$, such that $\partial\xi + \xi\partial = 1$. Let f and $f' : C' \rightarrow C$ be related by $f'f - 1 = \partial\eta + \eta\partial$, $ff' - 1 = \partial\eta' + \eta'\partial$, where $\eta : C \rightarrow C$, $\eta' : C' \rightarrow C'$ are deformation operators. Let

$$\mu = f\eta - \eta' f : C \rightarrow C'. \tag{8.6}$$

Then

$$\begin{aligned} \partial\mu + \mu\partial &= \partial f\eta - \partial\eta' f + f\eta\partial - \eta' f\partial = f(\partial\eta + \eta\partial) - (\partial\eta' + \eta'\partial)f \\ &= f(f'f - 1) - (ff' - 1)f = 0. \end{aligned}$$

Hence it follows by a straightforward calculation that $\partial\xi + \xi\partial = 1$, where $\xi : C(f) \rightarrow C(f)$ is given by

$$\begin{cases} \xi\alpha = \alpha\eta - \mu\eta \\ \xi \mid C' = \alpha f' - \mu f' - \eta'. \end{cases} \tag{8.7}$$

9. The groupoid $\mathfrak{G}$. Let R be the integral group ring of a group Γ. We need to consider chain mappings which do not necessarily commute with the operators $r \in R$. By a *chain mapping*, $f : C \rightarrow C'$, *associated with an automorphism*, $\theta : R \approx R$, we shall mean a family of homomorphisms, $f_n : C_n \rightarrow C'_n$, such that $f\partial = \partial f$ and $fr = (\theta r)f$. We now insist that $C_0 \neq 0$ and that, if m_i is a basis element of C_0, then $f_0 m_i$ shall be a basis element of C'_0. This ensures that f is associated with only one automorphism θ (unlike $C \rightarrow 0$, for example). If $f' : C' \rightarrow C''$ is associated with $\theta' : R \approx R$, then $f'f : C \rightarrow C''$ is obviously associated with $\theta'\theta$. Let $x \in R$ be any regular element. Then $\partial x = x\partial$ and $x(rc) = (xrx^{-1})xc$. Therefore $x : C \rightarrow C$, given by $c \rightarrow xc$, is a chain mapping associated with the inner automorphism θ_x. We shall confine ourselves to chain mappings associated with those automorphisms of R, which are determined by automorphisms of Γ, and we shall use the same symbol to denote $\theta : \Gamma \approx \Gamma$ and the corresponding automorphism of R.

We define chain homotopy and chain equivalence as in CH(II), with Γ playing the part of $\bar{\rho}_1$. Thus $f \simeq g : C \rightarrow C'$ means that

$$\gamma g - f = \partial\eta + \eta\partial, \tag{9.1}$$

where $\gamma \varepsilon \Gamma$ and $\eta : C \rightarrow C'$ is a chain deformation operator associated with the same automorphism, θ, as f. As in CH(II) it follows that g is associated with $\theta_\gamma^{-1}\theta$. We shall write $f \cong g$ if, and only if, f, g are related by (9. 1), with $\gamma = 1$. As in the ordinary theory of homotopy or chain homotopy, $f \simeq g$ implies $fh \simeq gh$, $h'f \simeq h'g$, where h, h' are any chain mappings of the form $h : C^0 \rightarrow C$, $h' : C' \rightarrow C''$.

We say that $f : C \rightarrow C'$ is a *chain equivalence* and write $f : C \equiv C'$, if, and only if there is a chain mapping, $g : C' \rightarrow C$, such that $gf \simeq 1$, $fg \simeq 1$. Let $\gamma, \gamma' \subset \Gamma$ be such that $\gamma gf \cong 1$, $\gamma' fg \cong 1$. Then $f'f \cong 1$, where $f' = \gamma g$. On transforming $\gamma' fg \cong 1$ by γ' we have $fg\gamma' \cong 1$. Therefore $ff' \cong ff'fg\gamma' \cong fg\gamma' \cong 1$.

Let $f : C \equiv C'$, where f is associated with $1 : R \approx R$. That is to say, $fr = rf$. Let $f' : C' \rightarrow C$ be such that $f'f \cong 1$, $ff' \cong 1$. Since $f'f$ is associated with only one $\theta : R \approx R$ and since $fr = rf$ it follows that $f'r = rf'$. Therefore $f : C \equiv C'$, in the sense of Lemma 6.

Let $f : C \equiv C'$ be associated with $\theta : R \approx R$ and let $\theta_1, \theta_2 : R \approx R$ be given. Using the same notation as in Section 7 we have

$$\partial'^{\theta_1} s_{\theta_1} f s_{\theta_2}{}^{-1} = s_{\theta_1} \partial' f s_{\theta_2}{}^{-1} = s_{\theta_1} f \partial s_{\theta_2}{}^{-1} = s_{\theta_1} f s_{\theta_2}{}^{-1} \partial^{\theta_2}.$$

Therefore

$$s_{\theta_1} f s_{\theta_2}{}^{-1} : C^{\theta_2} \rightarrow C'^{\theta_1} \tag{9.2}$$

is a chain mapping, which is obviously associated with $\theta_1\theta\theta_2{}^{-1}$. Let $f' : C' \rightarrow C$ be such that $f'f \cong 1$, $ff' \cong 1$. On transforming $f'f \cong 1$ by s_θ we have $s_\theta f' f s_\theta^{-1} \cong 1 : C^\theta \rightarrow C^\theta$. Also $f s_\theta^{-1} \cdot s_\theta f' \cong 1 : C' \rightarrow C'$. Therefore it follows from (9. 2), with $\theta_1 = 1$, $\theta_2 = \theta$, that $f s_\theta^{-1} : C^\theta \equiv C'$. Moreover $(f s_\theta^{-1}) r = r (f s_\theta^{-1})$. We define $\tau(f)$ by

$$\tau(f) = \tau(f s_\theta^{-1}) \tag{9.3}$$

and call it the *torsion* of f. Let $\theta_1, \theta_2 : R \approx R$ be arbitrary. Then $s_{\theta_1} f s_{\theta_2}{}^{-1}$ is associated with $\theta_1\theta\theta_2{}^{-1}$ and it follows from (9. 3) and (8. 5) that

$$\tau(s_{\theta_1} f s_{\theta_2}{}^{-1}) = \tau(s_{\theta_1} f s_{\theta_2}{}^{-1} \cdot s_{\theta_2} s_\theta^{-1} s_{\theta_1}{}^{-1}) = \tau(s_{\theta_1} \cdot f s_\theta^{-1} \cdot s_{\theta_1}{}^{-1}) = \theta_1 \tau(f). \tag{9.4}$$

Let $f' : C' \equiv C''$ and let f' be associated with $\theta' : R \approx R$. Then $f'f : C \equiv C''$ is associated with $\theta'\theta$, and it follows from Theorem 7 and (9.4) that

$$\begin{aligned} \tau(f'f) &= \tau(f'fs_\theta^{-1}s_{\theta'}^{-1}) = \tau(f's_{\theta'}^{-1} \cdot s_{\theta'}fs_\theta^{-1}\, s_{\theta'}^{-1}) \\ &= \tau(f') + \tau(s_{\theta'} \cdot fs_\theta^{-1} \cdot s_{\theta'}^{-1}) = \tau(f') + \theta'\tau(f). \end{aligned} \tag{9.5}$$

Let $f \cong g : C \equiv C'$ and let f, g be related by (9.1). Then $\gamma g, f$ and η are all associated with the same automorphism, θ, and

$$\gamma g s_\theta^{-1} - f s_\theta^{-1} = \partial'\eta s_\theta^{-1} + \eta s_\theta^{-1} s_\theta \partial s_\theta^{-1} = \partial'\xi + \xi\partial^\theta,$$

where $\xi = \eta s_\theta^{-1}$. Therefore $fs_\theta^{-1} \cong \gamma g s_\theta^{-1}$ and it follows from Theorem 8 that $\tau(f) = \tau(\gamma g)$. Also it follows from (9.5) and (2.11) that $\tau(\gamma g) = \tau(\gamma) + \theta_\gamma \tau(g) = \tau(\gamma) + \tau(g)$, where $\gamma : C' \to C'$ is the chain mapping $c' \to \gamma c'$, which is obviously a chain equivalence. Since $\gamma s_{\theta\gamma}^{-1} m_i = \gamma m_i$, where m_i is any basis element of C'_n, it follows from (2.7) that $\gamma s_{\theta\gamma}^{-1} : C'^{\theta\gamma} \approx C'\ (\Sigma)$, whence $\tau(\gamma) = 0$. Therefore $f \cong g$ implies

$$\tau(f) = \tau(g). \tag{9.6}$$

Let $\mathfrak{G}$ be the totality of chain homotopy classes of equivalences between all the chain systems, which are equivalent to a given one. Let $f : C \equiv C'$, $f' : C' \equiv C''$ be such equivalences and $\bar{f}$, $\bar{f}'$ the corresponding chain homotopy classes. We define $\bar{f}'\bar{f} = \overline{f'f}$. It may be verified that, when multiplication is thus defined, $\mathfrak{G}$ is a groupoid. Let f be associated with θ. Then we define $f : \mathrm{T} \to \mathrm{T}$ by $f\tau = \theta\tau$. Let $f \cong g$. Then g is associated with $\theta_\gamma^{-1}\theta$, for some $\gamma \varepsilon \Gamma$, and it follows from (2.11) that $g\tau = \theta_\gamma^{-1}\theta\tau = \theta\tau = f\tau$. Therefore a single-valued map $\bar{f}$: $\mathrm{T} \to \mathrm{T}$ is defined by $\bar{f}\tau = f\tau$. Obviously $1 \cdot \tau = \tau$ if 1 is any identical map, $C \to C$. Since $f'f$, if it exists, is associated with $\theta'\theta$, where f, f' are associated with θ, θ', it follows that $\bar{f}'(\bar{f}\tau) = (\bar{f}'\bar{f})\tau$. Therefore we say that T admits $\mathfrak{G}$ as a *groupoid of operators.*

It follows from (9.6) that a single-valued map $\tau : \mathfrak{G} \to \mathrm{T}$ is defined by $\tau(\bar{f}) = \tau(f)$ and from (9.5) that

$$\tau(\bar{g}\bar{f}) = \tau(\bar{g}) + \bar{g}\tau(\bar{f}). \tag{9.7}$$

Therefore τ is what, by a natural extension of the language of group theory, we call a *crossed homomorphism* of $\mathfrak{G}$ into T. We call $\tau(\bar{f})$ the *torsion* of $\bar{f}$. Given C it is easy to construct a chain system $C' \equiv C$ and an equivalence, $f : C \equiv C'$, such that $\tau(f)$ is a given element $\tau_0 \varepsilon \mathrm{T}$. For example, let $d : A \approx B$ be an isomorphism such that $\tau(d) = \tau_0$, where A, B are basic modules, which are disjoint from C and from each other Let $m > \dim C$ and let C' be the system which consists of C, with its own boundary operator, and

$C'_m = B$, $C'_{m+1} = A$, with $\partial'_{m+1} = d$. Then it follows from Theorem 10 that $\tau(i) = \tau_0$, where i is the identical map $C \to C'$.

10. Homotopy types of complexes.

Let K be a given complex and let a 0-cell $e^0 \varepsilon K^0$ be taken as base-point for $\pi_1(K)$. Let $\tilde{K}$ be the universal covering complex of K, in which the points are classes of paths joining e^0 to points in K. Let[26] $C(\tilde{K})$ be defined in the same way as $C(K)$ in Section **12** of CH(II) and let $(c^n{}_1, \cdots, c^n{}_{p_n})$ be a natural basis for $C_n(\tilde{K}) = \dot{H}_n(\tilde{K}^n, \tilde{K}^{n-1})$. Let R, Γ, M mean the same as before, with $\gamma : \pi_1(K) \approx \Gamma$. Let $R(K)$ be the group ring of $\pi_1(K)$. Let $C_n(K) \subset M$ be a basic module of rank p_n $(n = 0, 1, \cdots)$, such that $C_i(K) \cap C_j(K) = 0$ if $i \neq j$. Let $(m^n{}_1, \cdots, m^n{}_{p_n})$ be the basis of $C_n(K)$ and let $k_n : C_n(\tilde{K}) \approx C_n(K)$ be defined by

$$(10.1) \qquad k_n(r_1 c_{n1} + \cdots + r_{p_n} c^n{}_{p_n}) = (\gamma r_1) m^n{}_1 + \cdots + (\gamma r_{p_n}) m^n{}_{p_n},$$

where $r_i \varepsilon R(K)$. Let

$$(10.2) \qquad \partial_n = k_{n-1} \partial'_n k_n{}^{-1} : C_n(K) \to C_{n-1}(K),$$

where ∂' is the boundary operator in $C(\tilde{K})$. Obviously $\partial\partial = 0$ and $\partial r = r\partial$ $(r \varepsilon R)$. Therefore $C(K) = \{C_n(K)\}$, with $\partial = \{\partial_n\}$ as boundary operator, is a chain system and $k = \{k_n\} : C(\tilde{K}) \approx C(K)$ is an isomorphism associated with γ.

The arbitrariness in the definition of $C(K)$ consists of

a) the choice of the base point e^0,

(10.3) b) the choice of $\gamma : \pi_1(K) \approx \Gamma$,

c) the choices of the bases $\{c^n{}_i\}$ and of the basic modules $C_n(K)$.

Let another 0-cell $e_1{}^0 \varepsilon K^0$ be taken as base point and let $\tilde{K}_1$ and $C(\tilde{K}_1)$ be the corresponding universal covering complex and chain system. Let

$$\alpha : \pi_1(K, e_1{}^0) \approx \pi_1(K, e^0), \qquad \phi : \tilde{K}_1 \approx \tilde{K}$$

be the isomorphisms[27] determined by a path $(I, 0, 1) \to (K, e^0, e_1{}^0)$. Let $h : C(\tilde{K}_1) \approx C(\tilde{K})$ be the isomorphism induced by ϕ. Then h is obviously associated with α, and the same system, $C(K)$, is defined by

---

[26] Here we reserve the symbol $C(K)$ for a system in which $C_n(K)$ is a basic module in M, as defined by (10.1), (10.2) below. We no longer insist that the elementary chain $c^0 \varepsilon C_0(K)$, associated with a e-cell which covers e^0, shall be associated with the base point in K.

[27] We use the symbol $\approx$ to denote the relation of isomorphism between complexes as well as between groups and chain systems.

$$\gamma\alpha : \pi_1(K, e_1^0) \approx \Gamma, \qquad kh : C(\tilde{K}_1) \approx C(K),$$

as by α and k. The effect of choosing a different path, $(I, 0, 1) \rightarrow (K, e^0, e_1^0)$, is to replace α, h by $\theta_x\alpha$, xh, where $x \varepsilon \pi_1(K, e^0)$. Since

$$x(r_1c^n_1 + \cdots + r_{p_n}c^n_{p_n}) = (\theta_x r_1)xc^n_1 + \cdots + (\theta_x r_{p_n})xc^n_{p_n}$$

it follows that the resulting alterations, $\gamma\alpha \rightarrow \gamma\theta_x\alpha$ and $kh \rightarrow kxh$, are included in (b) and (c).

Let γ be replaced by $\gamma' : \pi_1(K) \approx \Gamma$ and let $\theta = \gamma'\gamma^{-1} : \Gamma \approx \Gamma$. Then it follows from (10.1) and (10.2) that k_n is replaced by $s_\theta k_n$ and ∂ by $\partial^\theta = s_\theta \partial s_\theta^{-1}$. Therefore $C(K)$ is replaced by the conjugate system $C^\theta(K)$.

Any other natural basis for $C_n(\tilde{K})$ is of the form $(\pm x_1c^n_1, \cdots, \pm x_{p_n}c^n_{p_n})$, where $x_i \varepsilon \pi_1(K)$. Any other basic module of rank p_n is of the form PC_n, where $P : M \rightarrow M$ is a permutation. Therefore a change in (c) leads to a new system $C'(K) \approx C(K)$ (Σ).

Therefore $C(K)$ is determined up to a transformation, $C(K) \rightarrow C'(K)$, which is the resultant of a semi-linear transformation, $C(K) \rightarrow C^\theta(K)$, followed by a simple isomorphism $C^\theta(K) \approx C'(K)$ (Σ).

Let $K' \equiv K$ and let $k' : C(\tilde{K}) \approx C(K')$ be defined in the same way as $k : C(\tilde{K}) \approx C(K)$, in terms of an isomorphism $\gamma' : \pi_1(K') \approx \Gamma$. Let $g : C(\tilde{K}) \equiv C(\tilde{K}')$ and $\alpha : \pi_1(K) \approx \pi_1(K')$ be the chain equivalence and the isomorphism induced by a homotopy equivalence[28] $\phi : K \equiv K'$. Let

$$(10.4) \qquad f = k'gk^{-1} : C(K) \rightarrow C(K'), \qquad \theta = \gamma'\alpha\gamma^{-1} : \Gamma \approx \Gamma.$$

Then it may be verified that f is a chain equivalence associated with θ. We describe it as the chain equivalence *induced* by ϕ and we define $\tau(\phi) = \tau(f)$. Let $g^* : C(\tilde{K}) \equiv C(\tilde{K}')$ be the chain equivalence induced by a homotopic map $\phi^* \simeq \phi$ and let $f^* = k'g^*k^{-1}$. Then $g^* \simeq g$ and it follows that $f^* \simeq f$. Therefore $\tau(\phi^*) = \tau(\phi)$. Hence, and by the two preceding paragraphs, $\tau(\phi)$ depends only on the homotopy class, $\bar{\phi}$, of maps $K \rightarrow K'$, which contains ϕ, on γ, γ' and on the choice of base points[29] in K, K'. We define $\tau(\bar{\phi}) = \tau(\phi)$. Let γ and γ' be replaced by $\gamma_1 : \pi_1(K) \approx \Gamma$ and $\gamma'_1 : \pi_1(K') \approx \Gamma$. Let $\theta = \gamma_1\gamma^{-1}$, $\theta' = \gamma'_1\gamma'^{-1}$. Then k, k' are replaced by $s_\theta k$, $s_{\theta'}k'$, and f by $s_{\theta'}fs_\theta^{-1}$. Therefore it follows from (9.4) that $\tau(\bar{\phi})$ is replaced by $\theta'\tau(\bar{\phi})$.

Let us write $\tau \equiv \tau'$, where $\tau, \tau \subset \mathrm{T}$, if, and only if, $\tau' = \theta\tau$, where

---

[28] All our maps and homotopies of complexes will be cellular and it is always to be understood that a given map, $K \rightarrow K'$, carries e^0 into e'^0, where $e^0 \varepsilon K'^0$ are the base points.

[29] Actually $\tau(\phi)$ does not depend on the choice of base points since $\theta_\gamma\tau = \tau$ for any $\gamma \varepsilon \Gamma$, $\tau \varepsilon \mathrm{T}$.

$\theta: \Gamma \approx \Gamma$. Obviously $\tau \equiv \tau'$ is an equivalence relation and we shall describe the corresponding equivalence classes as *θ-classes.* It follows from the preceding paragraph that the θ-class, $\bar{\tau}(\bar{\phi})$, which contains $\tau(\bar{\phi})$, is uniquely determined by the homotopy class $\bar{\phi}$. We call it the *torsion* of $\bar{\phi}$, or of any map $\phi \varepsilon \bar{\phi}$.

We shall describe $\phi: K \equiv K'$ as a *simple* (*homotopy*) *equivalence,* and shall write $\phi: K \equiv K'$ (Σ), if, and only if, $\tau(\phi) = 0$. We shall say that K, K' are of the same *simple homotopy type,* and shall write $K \equiv K'$ (Σ), if, and only if, there is a simple homotopy equivalence $\phi: K \equiv K'$ (Σ).

Let $\phi': K' \equiv K''$ and let $C(K'') \approx C(K'')$ be defined in the same way as $C(K)$ and $C(K')$, in terms of an isomorphism $\gamma'': \pi_1(K'') \approx \Gamma$. Let ϕ' be associated with $\alpha': \pi_1(K') \approx \pi_1(K'')$ and let $\theta' = \gamma''\alpha'\gamma'^{-1}: \Gamma \approx \Gamma$. Then it follows from (10. 4) and (9. 5) that

$$\tau(\phi'\phi) = \tau(\phi') + \theta'\tau(\phi). \tag{10. 5}$$

Therefore, if ϕ, ϕ' are simple equivalences, so is $\phi'\phi$. Obviously $\tau(\psi) = 0$ if $\psi \simeq 1: K \to K$. Therefore, taking $K'' = K$ and $\phi'\phi \simeq 1$, it follows from (10. 5) that a homotopy inverse of a simple homotopy equivalence is itself a simple homotopy equivalence. Therefore $K \equiv K'$ (Σ) is an equivalence relation.

Let G_K be the aggregate of homotopy classes, $\bar{\phi}, \bar{\psi}, \cdots$, of homotopy equivalences, $\phi, \psi, \cdots$, of K into itself. Let $\bar{\psi}\bar{\phi} = \overline{\psi\phi}$. Then G_K, with this multiplication, is obviously a group. Let $e^0 \varepsilon K^0$ and $k: C(K) \approx C(K)$ be fixed and let $\mathfrak{G}_K$ be the sub-group of the groupoid $\mathfrak{G}$, which consists of the chain homotopy classes of chain equivalences $C(K) \equiv C(K)$. Let $f_\phi = f$, where f is given by (10. 4), with $K' = K$, $\gamma' = \gamma$, $k' = k$, and let $\bar{f}_\phi \varepsilon \mathfrak{G}$ be the class which contains f_ϕ. Then $\bar{\phi} \to \bar{f}_\phi$ is obviously a homomorphism of G_K into $\mathfrak{G}_K$. Let $\tau_K: G_K \to \mathrm{T}$ be the map which is given by $\tau_K(\bar{\phi}) = \tau(\bar{\phi})$. It is the resultant of $\bar{\phi} \to \bar{f}_\phi$, followed by the crossed homomorphism $\bar{f} \to \tau(\bar{f})$. Therefore τ_K is a crossed homomorphism, in which G_K operates on T according to the rule $\bar{\phi}\tau = \bar{f}_\phi\tau$.

Let us take $\Gamma = \pi_1(K, p_0)$, where $p_0 \varepsilon K$, and let $\gamma: \pi_1(K, e^0) \approx \Gamma$ be the isomorphism determined by a path in K, which joins p_0 to e^0. Then the degree of arbitrariness in γ is that it may be replaced by $\theta\gamma$, where $\theta: \Gamma \approx \Gamma$ is an inner automorphism. In this case f_ϕ is replaced by $f_\phi{}^\theta$ and τ_K by $\theta\tau_K: G_K \to \mathrm{T}$. But $\theta\tau = \tau$, according to (2. 11), whence $\theta\tau_K = \tau_K$. Therefore τ_K is uniquely determined by K when $\Gamma = \pi_1(K, p_0)$. For the reasons given in discussing (10. 3), τ_K is independent of the choice of e^0.

Let K_0 be a connected sub-complex of K, which contains e^0 and is such that $i:\pi_1(K_0) \approx \pi_1(K)$, where i is the injection homomorphism. Then $\tilde{K}_0 = p^{-1}K_0$ may be taken as the universal covering complex of K_0, where $p:\tilde{K} \to K$ is the covering map. Let $C_n(\tilde{K}_0) \subset C_n(\tilde{K})$ be the sub-module consisting of the n-chains carried by $\tilde{K}_0$ and let k_n mean the same as in (10.1). A natural basis for $C_n(\tilde{K}_0)$ is part of a natural basis for $C_n(\tilde{K})$ and it follows that $C(K_0)$, with

$$C_n(K_0) = k_n C_n(\tilde{K}_0), \tag{10.6}$$

is a sub-system of $C(K)$. Let $U = K - K_0$ and let us denote the residue system $C(K) - C(K_0)$ by $C(U) = C(K) - C(K_0)$. Let $\tilde{U} = p^{-1}U$ and let $C_n(\tilde{U}) \subset C_n(\tilde{K})$ be the sub-module consisting of the n-chains carried by $\tilde{U}$. Then obviously $C_n(U) = k_n C_n(\tilde{U})$.

When dealing with such a pair of complexes K and $K_0 \subset K$ we shall always assume that $C(K_0)$ is imbedded in $C(K)$ in the way described above.

Let $K_0 \subset K$, $L_0 \subset L$ be sub-complexes of given complexes K, L. Let $\phi: (K, K_0) \to (L, L_0)$ be a map such that $\phi \mid K - K_0$ is an isomorphism onto $L - L_0$ and let $\phi_0: K_0 \to L_0$ be the map which is induced by ϕ.

THEOREM [30] 11. *If ϕ_0 is a simple equivalence, so is ϕ.*

Let $h: C(K) \to C(L)$ and $f: C(K_0) \to C(L_0)$ be the chain mappings which are induced by ϕ and ϕ_0. Then it is obvious that $hC(K_0) \subset C(L_0)$ and that f is the chain mapping induced by h. Since $\phi \mid K - K_0$ is an isomorphism onto $L - L_0$ it is also obvious that $g: C(K - K_0) \approx C(L - L_0)$ (Σ), where g is the chain mapping induced by h. Therefore the Theorem follows from Theorem 10.

As an application of Theorem 11 let $\phi_0: K_0 \equiv L_0$, where L_0 consists of a single 0-cell, let L be formed from K by shrinking K_0 into the point L_0 and let $\phi: K \to L$ be the "identification map." Since $\pi_1(L_0) = 1$ it follows that ϕ_0 is a simple equivalence and so therefore is ϕ. In particular we can take $K_0 \subset K^1$ to be a tree containing K^0. Then L^0 consists of the single 0-cell L_0.

11. Combinatorial invariance.

In this section we prove:

THEOREM 12. *If K' is a sub-division of K the identical map, $i: K \to K'$, is a simple equivalence.*

---

[30] Cf. Theorem 12 in Section 8 of (I). Obviously $\phi_0: K_0 \equiv L_0(\Sigma)$ if $K_0 = e^0$, $L_0 = e'^0$, where $e^0 \varepsilon K^0$, $e'^0 \varepsilon L^0$ are the base points. In this case the theorem states that $\phi: K \equiv L(\Sigma)$ if $\phi: K \approx L$.

Let P be a given complex and $Q \subset P$ a sub-complex, which is a D. R. of P.

LEMMA 10. *If every circuit in $P - Q$ is contractible to a point in P, then the identical map, $Q \rightarrow P$, is a simple equivalence.*

We first prove the theorem, assuming the truth of the lemma. Let $\phi' : K' \approx L$, where L is a new complex, which does not meet K. Let $\phi = \phi' i : K \rightarrow L$. Then $i = \phi'^{-1}\phi$ and $\phi'^{-1} : L \equiv K'$ (Σ) according to Theorem 11. Therefore it is sufficient to prove that $\phi : K \equiv L$ (Σ). Let P be the mapping cylinder of ϕ. We regard P as $K \times I$, with $(x, 0) = x$ $(x \varepsilon K)$ and $K \times 1$ sub-divided to form L. Let e_0 be a principal cell (i. e. one which is an open sub-set) of K and let $K_1 = K - e_0$. Proceeding by induction we define a sequence of sub-complexes

$$K = K_0, K_1, \cdots, K_n = K^{-1},$$

such that $K_{\lambda+1} = K_\lambda - e_\lambda$, where e_λ is a principal cell of K_λ. Let $P_\lambda = K \cup (K_\lambda \times I)$. Then $P_{\lambda+1}$ is a D. R.[31] of P_λ and $P_\lambda - P_{\lambda+1}$ is the point-set $e_\lambda \times (0, 1\rangle$, where $(0, 1\rangle$ is the half open interval $0 < t \leq 1$. Therefore $\pi_1(P_\lambda - P_{\lambda+1}) = 1$ and it follows from .Lemma 10 that $i_\lambda : P_{\lambda+1} \equiv P_\lambda$ (Σ), where i_λ is the identical map. Therefore

$$j = i_0 \cdots i_{n-1} : P_n = K \equiv P \ (\Sigma).$$

Similarly $k : L \equiv P$ (Σ), where k is the identical map. Let $\psi : P \rightarrow L$ be given by $\psi(x, t) = \phi x$. Then $\psi : P \equiv L$ (Σ), since ψ is a homotopy inverse of k. Therefore $\phi = \psi j : K \equiv L$ (Σ) and Theorem 12 is proved.

It remains to prove Lemma 10. Since Q is a D. R. of P it is easily proved that the chain system $C(Q)$ is a D. R. of $C(P)$ and that a retraction $\psi : P \rightarrow Q$ induces a retraction $k : C(P) \rightarrow C(Q)$. We have to prove that k is a simple equivalence and this will follow from Theorem 4, Section **5**, when we have proved that $C(U) \equiv 0$ (Σ), where $U = P - Q$.

Let $U_1, \cdots, U_m$ be the components of U, which are finite in number since U is the union of a finite number of (connected) cells. Let $\tilde{P}$ be the universal covering complex of P and let $\mathbf{p} : P \rightarrow \tilde{P}$ be the covering map. Let $\tilde{U}_\lambda$ be any component of $\mathbf{p}^{-1}U_\lambda$ and let $U^* = \tilde{U}_1 \cup \cdots \cup \tilde{U}_m$. Since P and $\tilde{P}$ are locally connected, U_λ and $\tilde{U}_\lambda$ are open sets. Let $e^n{}_1, \cdots, e^n{}_{q_n}$ be the n-cells in U. It follows from the condition on the circuits in U, which is satisfied a fortiori by the circuits in U_λ, that $\mathbf{p} \mid \tilde{U}_\lambda$ is a homeomorphism onto U_λ. Therefore $\mathbf{p} \mid U^*$ is a homeomorphism onto U. Therefore

---

[31] See Theorem 1.4(ii) in [16].

U^* contains precisely one, $\tilde{e}^n_i$, of the cells in $\bar{P}$, which cover e^n_i. Let $c^n_i \varepsilon C_n(\bar{P})$ be the element which is represented by a characteristic map for $\tilde{e}^n_i$. Then $(c^n_1, \cdots, c^n_{q_n})$ is a basis for $C_n(\tilde{U})$, which is part of a natural basis for $C_n(\bar{P})$. Moreover $C_n(U^*)$, which consists of the n-chains carried by U^*, is the ordinary free Abelian group, which is freely generated by $c^n_1, \cdots, c^n_{q_n}$, without the help of the operators in $\pi_1(P)$.

Since each component of $\tilde{U}$ is open it follows that no cell in $\tilde{U} - U^*$ meets the closure of $\tilde{e}^n_i$. Therefore

$$\partial c^n_i = \sum_{j=1}^{q_{n-1}} d_{ij}{}^n c_j{}^{n-1} + c'_i{}^{n-1}, \tag{11.1}$$

where $d_{ij}{}^n$ are integers and $c'_i{}^{n-1} \varepsilon C_n(\tilde{Q})$ $(\tilde{Q} = p^{-1}Q)$. Let $m_i{}^n = k_n c_i{}^n$, where k_n means the same as in (10.1). Then it follows from (11.1) that $\partial : C(U) \rightarrow C(U)$ is given by $\partial m_i{}^n = \sum_{j=1}^{q_{n-1}} d_{ij}{}^n m_j{}^{n-1}$. Therefore $C(U) \equiv 0\ (\Sigma)$, by the corollary to Lemma 4, in Section **6**. This proves Lemma 10.

12. Lens spaces. By way of an example let A, B be the chain systems determined by lens spaces of types (m, p), (m, q), where m is the order of their fundamental groups and $q \equiv k^2 p(m)$. That is to say, A, B play the part of $C(K)$ in Section **10** and $\pi_1(P)$, $\pi_1(Q)$ in Section **15** of CH(II) are both replaced by Γ. The generators ξ, η in CH(II) are replaced by a generator $\gamma \varepsilon \Gamma$ and we denote the integer r by h, to avoid confusion with $r \varepsilon R$. Otherwise the notations will be the same as in CH(II). Thus $\partial : A \rightarrow A$ and $\partial : B \rightarrow B$ are given by

$$\begin{cases} \partial a_1 = (\gamma - 1) a_0, & \partial a_2 = \sigma_m(\gamma) a, \quad \partial a_3 = (\gamma^p - 1) a_2 \\ \partial b_1 = (\gamma - 1) b_0, & \partial b_2 = \sigma_m(\gamma) b_1, \quad \partial b_3 = (\gamma^q - 1) b_2, \end{cases} \tag{12.1}$$

where $\sigma_t(\gamma^s) = 1 + \gamma^s + \cdots + \gamma^{(t-1)s}$.

The Reidemeister-Franz torsion in A and B is r_A and r_B, where

$$r_A = (\gamma - 1)(\gamma^p - 1), \qquad r_B = (\gamma - 1)(\gamma^q - 1). \tag{12.2}$$

Let $\theta : \Gamma \approx \Gamma$ be given by $\theta\gamma = \gamma^k$ and let $u : A \rightarrow B$, $v : B \rightarrow A$ be the chain mappings, associated with θ and with θ^{-1}, which are given by $ua_n = b_n$, $vb_n = a_n$ if $n = 0$ or 3 and

$$\begin{cases} ua_1 = \sigma_k(\gamma) b_1, & ua_2 = \sigma_k(\gamma^{kp}) b_2 \\ vb_1 = \sigma_l(\gamma) a_1, & vb_2 = \sigma_l(\gamma^{lq}) a_2, \end{cases} \tag{12.3}$$

where $kl = 1 + mh$. As shown in CH(II), $vu - 1 = \partial\eta + \eta\partial$, $uv - 1$

$= \partial\eta + \eta\partial$, where $\eta a_n = \eta b_n = 0$ if $n \neq 1$ and $\eta a_1 = ha_2$, $\eta b_1 = hb_2$. Notice that $\eta\eta = 0$ and $\eta s_\theta = s_\theta\eta$, since a_n, b_n are generators, m_{i_n}, m_{j_n}, of M.

It follows from (12. 1) that

$$(12.\ 4) \qquad \partial^\theta a_1 = (\gamma^k - 1)a_0, \qquad \partial^\theta a_2 = \sigma_m(\gamma)a_1, \qquad \partial^\theta a_3 = (\gamma^{kp} - 1)a_2,$$

and from (12. 3) that

$$(12.\ 5) \qquad s_\theta v b_1 = \sigma_l(\gamma^k)a_1, \qquad s_\theta v b_2 = \sigma_l(\gamma^q)a_2,$$

since $kl \equiv 1(m)$. We proceed to calculate $\tau(u) = \tau\{C(us_\theta^{-1})\}$, where $C(f)$ means the same as in Section 8. Let $C = C(us_\theta^{-1})$ and let $a'_n = \alpha a_n$, where α means the same as in (8. 2). Then (b_n, a'_{n-1}) is a basis for C_n $(a'_{-1} = b_4 = 0)$. Since $us_\theta^{-1}a_n = ua_n$ it follows from (8. 2), (12. 3) and (12. 4) that $\partial : C \to C$ is given by

$$\begin{array}{ll}
\partial b_1 = (\gamma - 1)b_0, & \partial a'_0 = b_0 \\
\partial b_2 = \sigma_m(\gamma)b_1, & \partial a'_1 = \sigma_k(\gamma)b_1 - (\gamma^k - 1)a'_0 \\
\partial b_3 = (\gamma^q - 1)b_2, & \partial a'_2 = \sigma_k(\gamma^{kp})b_2 - \sigma_m(\gamma)a'_1 \\
 & \partial a'_3 = b_3 - (\gamma^{kp} - 1)a'_2.
\end{array}$$

It is easily verified that $\mu = 0$, where μ is given by (8. 6), and similarly that $s_\theta v\eta = \eta s_\theta v$. Since $\eta\eta = 0$ a straightforward caluculation shows that $\delta\delta = 0$, where δ is given by (8. 7), with $\xi = \delta$. It follows from (8. 7) and (12. 5) that $\delta b_0 = a'_0$ and

$$\begin{array}{ll}
\delta b_1 = -hb_2 + \sigma_l(\gamma^k)a'_1, & \delta a'_0 = 0 \\
\delta b_2 = \sigma_l(\gamma^q)a'_2, & \delta a'_1 = ha'_2 \\
\delta b_3 = a'_3, & \delta a'_2 = 0.
\end{array}$$

Let $D_0 = C_0 + C_2 + C_4$, $D_1 = C_1 + C_3$, as in (6. 8) with $m = \dim C - 1 = 3$. Then D_0, D_1 have (b_0, a'_1, b_2, a'_3), (a'_0, b_1, a'_2, b_3) as bases and $\Delta_0 = \partial + \delta : D_0 \to D_1$ is given by

$$\begin{array}{lll}
\Delta_0 b_0 = a'_0 & & \\
\Delta_0 a'_1 = -(\gamma^k - 1)a'_0 + & \sigma_k(\gamma)b_1 + ha'_2 & \\
\Delta_0 b_2 = & \sigma_m(\gamma)b_1 + \sigma_l(\gamma^q)a'_2 & \\
\Delta_0 a'_3 = & & -(\gamma^{kp} - 1)a'_2 + b_3.
\end{array}$$

Let $f_0 : D_0 \approx D_0$, $f_1 : D_1 \approx D_1$ be the simple automorphisms which are given by

$$\begin{array}{lll}
f_0 a'_1 = a'_1 + (\gamma^k - 1)b_0, & f_0 c = c & (c = b_0, b_2, a'_3) \\
f_1 b_3 = b_3 + (\gamma^{kp} - 1)a'_2, & f_1 c = c & (c = a'_0, b_1, a'_2).
\end{array}$$

Then $\Delta'_0 = f_1\Delta_0 f_0 : D_0 \rightarrow D_1$ is given by $\Delta'_0 b_0 = a'_0$, $\Delta'_0 a'_3 = b_3$ and

$$\Delta'_0 a'_1 = \sigma_k(\gamma) b_1 + h a'_2$$
$$\Delta'_0 b_2 = \sigma_m(\gamma) b_1 + \sigma_l(\gamma^q) a'_2.$$

Let i, j be any integers and let $\rho = (i, j)$. If $i \leq j$ we have, writing $\sigma_s = \sigma_s(\gamma)$ $(\sigma_0 = 0)$,

$$\sigma_j - \gamma^{j-i}\sigma_i = \sigma_{j-i}.$$

Therefore it follows by induction on $i + j$ that the matrix $[\sigma_i, \sigma_j]$ can be reduced to $[\sigma_\rho, 0]$ by a sequence of transformations of the form

$$[\sigma_i, \sigma_j] \rightarrow [\sigma_i, \sigma_j - \gamma^s\sigma_i] \text{ or } [\sigma_i - \gamma^s\sigma_j, \sigma_j]$$

followed, if necessary, by $[0, \sigma_\rho] \rightarrow [\sigma_\rho, 0]$. Since $(k, m) = 1$ it follows that

$$\begin{bmatrix} \sigma_k(\gamma) & h \\ \sigma_m(\gamma) & \sigma_l(\gamma^q) \end{bmatrix} \rightarrow \begin{bmatrix} 1 & r \\ 0 & t \end{bmatrix} \qquad (r, t \subset R)$$

by such elementary transformations of the rows. Since these transformations alter the determinant by a factor ± 1, at most, we have $t = \pm d$, where

$$d = \sigma_k(\gamma)\sigma_l(\gamma^q) - h\sigma_m(\gamma).$$

I say that

$$(12.6) \qquad dr_B = \theta r_A,$$

where r_A, r_B are given by (12.2). For let χ be the homomorphism of R into the complex field, which is given by $\chi r = \omega$, where $\omega^m = 1$. Then $\chi(dr_B) = \chi(\theta r_A) = 0$ if $\omega = 1$, and if $\omega \neq 1$ we have

$$\chi(dr_B) = [(\omega^k - 1)(\omega^{lq} - 1)]/[(\omega - 1)(\omega^q - 1)](\omega - 1)(\omega^q - 1)$$
$$= (\omega^k - 1)(\omega^{lq} - 1) = \chi(\theta r_A).$$

Therefore $\chi(dr_B - r_A) = 0$ and (12.6) follows from the orthogonality relations between the group characters. Therefore $\tau(u)$ is essentially the same as the inverse of the element π in Lemma 5 on p. 1209 of [3].

13. Formal deformations. As in CH(II) let I^n be the n-cube in Hilbert space, which is given by $0 \leq t_1, \cdots, t_n \leq 1$, $t_i = 0$ if $i > n$, with $I^0 = (0, 0, \cdots)$. Let

$$E_0^{n-1} = \partial I^n - (I^{n-1} - \partial I^{n-1}) \qquad (n \geq 1).$$

Let e^n be a principal cell of a complex K and let e^{n-1} be a principal cell of $K - e^n$ $(n \geqq 1)$. We shall describe the transformation $K \rightarrow K_1 = K - e^n - e^{n-1}$ as an *elementary contraction* if, and only if, e^n has a characteristic map, $f: I^n \rightarrow \bar{e}^n$, such that $fE_0^{n-1} \subset K_1$ and $f \mid I^{n-1}$ is a characteristic map for e^{n-1}. The inverse, $K_1 \rightarrow K$, of an elementary contraction, $K \rightarrow K_1$, will be called an *elementary expansion.* An elementary expansion may also be defined as follows. Let E^n be an n-element, which is disjoint from a given complex, K, and let E^{n-1} be a hemisphere[32] of ∂E^n. Let $e^n = E^n - \partial E^n$, $e^{n-1} = \partial E^n - E^{n-1}$ and let $f: (E^{n-1}, \partial E^{n-1}) \rightarrow (K^{n-1}, K^{n-2})$ be an arbitrary map. Let $K_1 = K \cup e^{n-1} \cup e^n$ be the complex formed by identifying each point $p \varepsilon E^{n-1}$ with $fp \varepsilon K^{n-1}$. Then $K \rightarrow K_1$ is obviously an elementary expansion.

Either an elementary expansion or an elementary contraction will be called an *elementary deformation.* The resultant, $K_0 \rightarrow K_r$, of a finite sequence of elementary deformations,

$$K_i \rightarrow K_{i+1} \qquad (i = 0, \cdots, r-1), \tag{13.1}$$

will be called a *formal deformation.* We also include the identical transformation, $K \rightarrow K$, of any complex, among the formal deformations. We shall denote a formal deformation by $D: K_0 \rightarrow K_r$, and $K_r = DK_0$ will mean that K_r is obtained from K_0 by a formal deformation D. If each $K_i \rightarrow K_{i+1}$ is an elementary expansion (contraction) then $K_0 \rightarrow K_r$ will be called an *expansion* (*contraction*) and we shall say that K_0 *expands* (*contracts*) into K_r. If $D: K_0 \rightarrow K_r$ is the resultant of the sequence (13.1), then the resultant of the sequence $K_{i+1} \rightarrow K_i$ is the formal deformation, $D^{-1}: K_r \rightarrow K_0$, *inverse* to D.

Let $D: K_0 \rightarrow K_1$ be an elementary contraction. Obviously E_0^{n-1} is a D.R. of I^n. Therefore[33] K_1 is a D.R. of K_0 and $[D]$ will denote the homotopy class of maps, $K_0 \rightarrow K_1$, which contains a retraction. If $D: K_0 \rightarrow K_1$ is an elementary expansion then $[D]$ will denote the homotopy class of maps, $K_0 \rightarrow K_1$, which contains the identity. Let $D = D_{r-1} \cdots D_0: K_0 \rightarrow K_r$, where D_i stands for (13.1). We define $[D]$ by $[D] = [D_{r-1}] \cdots [D_0]$. If $1: K \rightarrow K$ is the identical formal deformation then $[1]$ will denote the homotopy class of the identical map $K \rightarrow K$. It is obvious that, if $D: K \rightarrow K'$ and $D': K' \rightarrow K''$ are formal deformations, then $K'' = D'DK$ and

$$[D'D] = [D'][D]. \tag{13.2}$$

Also it is easily verified that $[D^{-1}][D] = [1]$.

---

[32] I. e. the image of E_0^{n-1} in some homeomorphism $\partial I^n \rightarrow \partial E^n$.

[33] See Lemma 2 in Section 4 of [4].

Let L be a sub-complex of K, which may be empty. By a formal deformation, $D: K \to K'$, rel. L, we shall mean the resultant of a sequence of elementary deformations, none of which removes a cell of L. By $K' = DK$, rel. L, we shall mean that K' is the image of K in such a formal deformation, D. If $K' = DK$, rel. L, then $[D]$ obviously contains at least one map $\phi_0: K \to K'$, rel. L, where rel. L means that $\phi_0 y = y$ if $y \varepsilon L$. We restrict $[D]$ to maps, $\phi: K \to K'$, rel. L, such that $\phi \simeq \phi_0$, rel. L.

We shall describe a map $\phi: K \to K'$, rel. L, as a *restricted equivalence*, rel. L, if, and only if, $K' = DK$, rel. L, and $\phi \varepsilon [D]$. We shall write $K \equiv K' (\Sigma)$, rel. L, if, and only if, there is a simple equivalence $\phi: K \equiv K' (\Sigma)$, which is rel. L.

THEOREM 13. *$K' = DK$, rel. L, if, and only if, $K \equiv K' (\Sigma)$, rel. L, in which case the restricted equivalences, $K \to K'$, rel. L, are the same as the simple equivalences, $K \to K'$, rel. L.*

In order to prove this we shall need some lemmas, which are proved in the following section.

14. Lemmas on formal deformations. Let K, K' be complexes, with a common sub-complex, L, which may be empty, and let $(K - L) \cap K' = 0$. Let $\phi: K \to K'$ be a map rel. L. By the *mapping cylinder* of ϕ we mean the complex, P, which is formed from[34] $K \times I$ by identifying $(x, 0)$ with x, $(x, 1)$ with ϕx and $y \times I$ with y, for each point $x \varepsilon K$ and $y \varepsilon L$. Let

$$\phi_r: \pi_r(K) \approx \pi_r(L) \qquad (r = 1, \cdots, m), \tag{14.1}$$

where ϕ_r is the homomorphism induced by ϕ. Then the argument[35] used in the case $L = 0$ shows that

$$i_r: \pi_r(K) \approx \pi_r(P) \qquad (r = 1, \cdots, m), \tag{14.2}$$

where i_r is the injection. Therefore it follows from the exactness of the sequence

$$\pi_r(K) \to \pi_r(P) \to \pi_r(P, K) \to \pi_{r-1}(K) \to \pi_{r-1}(P),$$

that $\pi_r(P, K) = 0$ for $r = 1, \cdots, m$, where $\pi_1(P, K) = 0$ means that i_1 is onto $\pi_1(P)$.

---

[34] After replacing $K \times I$ by a homeomorph, if necessary, we assume that it has no point in common with K or K'.

[35] See Section 3 of [5].

LEMMA 11. *P contracts into K'.*

Let $K = K_0 \cup e^n$, where e^n is a principal cell in $K - L$. Then $P = P_0 \cup e^n \cup e^{n+1}$, where $e^{n+1} = e^n \times (0,1)$ and P_0 is the mapping cylinder of $\phi \mid K_0 : K_0 \rightarrow K'$. Let $f : I^n \rightarrow \bar{e}^n$ be a characteristic map for e^n. Then $g : I^{n+1} \rightarrow \bar{e}^{n+1}$, given by $g(t_1, \cdots, t_n, t) = \{f(t_1, \cdots, t_n), t\}$, is obviously a characteristic map for e^{n+1} and $gE_0{}^n \subset P_0$ and $gx = fx$ if $x \varepsilon I^n$. Therefore P contracts into P_0. Therefore the lemma follows by induction on the number of cells in $K - L$.

Let $\psi : P \rightarrow K'$ be given by

$$\psi \mid K' = 1, \qquad \psi(x, t) = \phi x \qquad\qquad (x \varepsilon K).$$

Since $\psi \mid K' = 1$ and since any two retractions $P \rightarrow K'$ are homotopic to each other, we have $\psi \varepsilon [D]$, where $D : P \rightarrow K'$ is any contraction. Also $\phi = \psi i$, where $i : K \rightarrow P$ is the identical map. If $K = D_1 P$, rel. K, then $i \varepsilon [D_1{}^{-1}]$. Hence, and from (13.2), we have the corollary:

COROLLARY. *If $K = D_1 P$, rel. K, then $\phi \varepsilon [DD_1{}^{-1}]$.*

Let K, K' and L be as in Lemma 11, except that $K - L$ and $K' - L$ may now have points in common. Let $\phi : (K, L) \approx (K', L)$, rel. L.

LEMMA 12. *ϕ is a restricted equivalence, rel. L.*

First let $K \cap K' = L$ and let P be the mapping cylinder of ϕ. Then P may also be regarded as the mapping cylinder of ϕ^{-1} and the Lemma follows from Lemma 11 and its corollary.

If $K' - L$ meets K we replace the points in $K' - L$ by new ones, thus forming a complex K'', such that $\phi' : K'' \approx K'$, rel. L, and $K \cap K'' = K' \cap K'' = L$. By what we have already proved, ϕ' and $\phi'^{-1}\phi : K \approx K''$ are restricted equivalences, rel. L. Therefore it follows from (13.2) that ϕ is a restricted equivalence, rel. L, and the lemma is proved.

Let K_0, K_1 be complexes with a common sub-complex, K, and let $K_i = K \cup e_i{}^n$ $(i = 0, 1; n \geq 1)$. Let $f_i : I^n \rightarrow e_i{}^n$ be a characteristic map for $e_i{}^n$ in K_i.

LEMMA 13. *If $f_0 \mid \partial I^n \simeq f_1 \mid \partial I^n$ in K, then $K_1 = DK_0$, rel. K.*

First assume that $e_0{}^n \cap e_1{}^n = 0$ and unite K_0, K_1 in the complex $K^* = K_0 \cup K_1$. Let $g_t : \partial I^n \rightarrow K$ be a homotopy of $g_0 = f_0 \mid \partial I^n$ into $g_1 = f_1 \mid \partial I^n$ and let $f : \partial I^{n+1} \rightarrow K^*$ be given by

$$f(t_1,\cdots,t_n,i)=f_i(t_1,\cdots,t_n) \qquad \{i=0,1;\ (t_1,\cdots,t_n)\,\varepsilon\, I^n\}$$
$$f(s_1,\cdots,s_n,t)=g_t(s_1,\cdots,s_n) \qquad \{t\,\varepsilon\, I;\ (s_1,\cdots,s_n)\,\varepsilon\,\partial I^n\}.$$

We attach a new cell, e^{n+1}, to K^* by means of the map[36] f, thus forming a complex, $L=K^*\cup e^{n+1}$, in which e^{n+1} has a characteristic map, $h: I^{n+1}\rightarrow \bar{e}^{n+1}$ such that $h\mid \partial I^{n+1}=f$. Since $h(x,0)=f_0x$ $(x\,\varepsilon\, I^n)$ and[37] $hE_0{}^n\subset L-e_0{}^n$ it follows that $L\rightarrow K_1$ is an elementary contraction. Similarly $L\rightarrow K_0$ is an elementary contraction. Therefore $K_0\rightarrow L\rightarrow K_1$ is a formal deformation, rel. K.

If $e_0{}^n\cap e_1{}^n\neq 0$ we attach a new cell, $e'_0{}^n$, to K, by means of the map $f_0\mid\partial I^n$, taking care that $e'_i{}^n\cap e_i{}^n=0$ $(i=0,1)$. Then $K_0\rightarrow K\cup e'_0\rightarrow K_1$ is a formal deformation, rel. K, and the lemma is proved.

Let P be a given complex, let $P_0\subset P$ be a sub-complex and let $k_n(P-P_0)$ denote the number of n-cells in $P-P_0$. Let K be a sub-complex of P_0 and let $D_0: P_0\rightarrow Q_0$ be a formal deformation, rel. K. We shall describe a formal deformation, $D: P\rightarrow Q$, rel. K, as an *extension* of D_0 if, and only if, Q_0 is a sub-complex of Q and

$$k_n(Q-Q_0)=k_n(P-P_0) \qquad (n=0,1,\cdots).$$

LEMMA 14. *$D_0: P_0\rightarrow Q_0$ has an extension $D: P\rightarrow Q$.*

Let $D: P\rightarrow Q$ be an extension of D_0 and let $D': Q\rightarrow Q'$, rel. K, be an extension of a formal deformation $D'_0: Q_0\rightarrow Q'_0$. Then $D'D: P\rightarrow Q'$ is obviously an extension of $D'_0D_0: P_0\rightarrow Q'_0$. Let $P_1\subset P$ be a sub-complex, which contains P_0. Let $D_1: P_1\rightarrow Q_1$, rel. K, be an extension of D_0 and let $D: P\rightarrow Q$ be an extension of D_1. Then D is obviously an extension of D_0. Therefore the Lemma will follow by a double induction on the number of elementary deformations in D_0 and on the number of cells in $P-P_0$ when we have proved it in case D_0 is an elementary deformation and $P-P_0$ is a single cell.

Let $P=P_0\cup e^n$ and let D_0 be an elementary expansion $D_0: P_0\rightarrow Q_0=P_0\cup e^{p-1}\cup e^p$. If e^n has a point in common with $e^{p-1}\cup e^p$ we apply a preliminary formal deformation, $P\rightarrow P'$, rel. P_0, as in Lemma 13, so as to replace e^n by a cell which is disjoint from $e^{p-1}\cup e^p$. Then P' and Q_0 may be united in a complex

$$Q=P'\cup e^{p-1}\cup e^p$$

---

[36] That is to say, we attach an $(n+1)$-element, E^{n+1} $(e^{n+1}=E^{n+1}-\partial E^{n+1})$, to K^* by means of the map $fh': \partial E^{n+1}\rightarrow K^*$, where $h': \partial E^{n+1}\rightarrow\partial I^{n+1}$ is a homomorphism.

[37] We recall that $E_0{}^{n+1}=\partial I^{n+1}-(I^n-\partial I^n)$.

and $P \to P' \to Q$ is an extension of D_0.

Let D_0 be an elementary contraction,

$$D_0 : P_0 \to Q_0 = P_0 - e^p - e^{p-1},$$

and let $f : I^n \to \bar{e}^n$ be a characteristic map for e^n. Since Q_0 is a D. R. of P_0 there is a map, $f' : \partial I^n \to Q_0$, which is homotopic, in P_0, to $f \mid \partial I^n$. We attach a new cell, e'^n, to P_0 by means of the map f', thus forming a complex $P' = P_0 \cup e'^n$. Then $P' = D'P$, rel. P_0, by Lemma 13. Since $\partial e'^n \subset Q_0$ it follows that $P' \to Q = P' - e^p - e^{p-1}$ is an elementary contraction and $P \to P' \to Q$ is an extension of D_0. This proves the lemma.

Let K be a (connected) sub-complex of P such that $\pi_n(P, K) = 0$ $(n = 1, \cdots, r)$.

LEMMA 15. *There is a formal deformation $D : P \to Q$, rel. K, such that $k_n(Q - K) = 0$ if $n \leq r$ and $k_n(Q - K) = k_n(P - K)$ if $n > r + 2$.*

Let $0 \leq p \leq r$ and assume that, if $p > 0$, then $k_n(P - K) = 0$ for $n = 0, \cdots, p - 1$. For the sake of clarity we consider the case $p = 0$ separately. Let $p = 0$ and let $e_1^0, \cdots, e_k^0$ be the 0-cells in $P - K$. Since P is connected there is a map

$$g_i : (E_0^1, I^0, E_0^0) \to (P^1, e_i^0, K^0).$$

Let $E_1^2, \cdots, E_k^2$ be a set of 2-elements, which are disjoint from P and from each other. Let $h_i : \partial E_i^2 \to \partial I^2$ be a homeomorphism and let

$$e_i^2 = E_i^2 - \partial E_i^2, \qquad e_i^1 = h_i^{-1}(I^1 - \partial I^1) = h_i^{-1}(\partial I^2 - E_0^1).$$

We attach E_i^2 to P by means of the map $g_i h_i : h_i^{-1} E_0^1 \to P^1$, thus forming a complex $P^* = P \cup \bigcup_i (e_i^1 \cup e_i^2)$. Then $P \to P^*$ is an expansion. The complex $K^* = K \cup \bigcup_i (e_i^0 \cup e_i^1)$ contracts into K. By Lemma 14 there is an extension, $P^* \to M_0$, rel. K, of the contraction $K^* \to K$. Then

$$k_0(M_0 - K) = k_0(P^* - K^*) = 0$$

$$k_n(M_0 - K) = k_n(P^* - K^*) = k_n(P - K) \qquad (n > 2).$$

Thus we have eliminated the 0-cells from $P - K$ at the expense of introducing $k_0(P - K)$ new 2-cells.

Now let $p > 0$, let $e_1^p, \cdots, e_k^p$ be the p-cells in $P - K$ and let $f_i : I^p \to \bar{e}_i^p$ be a characteristic map for e_i^p. Since $k_{p-1}(P - K) = 0$, whence

$P^{p-1} = K^{p-1}$, we have $f_i \partial I^p \subset K$. Since $\pi_p(P, K) = 0$ and since I^p, $E_0{}^p$ are two hemispheres of $\partial E_0{}^{p+1} = \partial I^{p+1}$, the map f_i can be extended to a map,

$$g_i\colon (E_0{}^{p+1}, E_0{}^p) \to (P^{p+1}, K^p). \tag{14.3}$$

It now follows, in exactly the same way as when $p = 0$, that there is a complex $M_p = D_p P$, rel. K, such that $k_n(M_p - K) = 0$ if $n \leq p$ and $k_n(M_p - K) = k_n(P - K)$ if $n > p + 2$. Therefore the lemma follows by induction on p.

15. Proof of Theorem 13. Let $D\colon K \to K_1 = K \cup e^{n-1} \cup e^n$, $(n \geq 1)$ be an elementary expansion. Then $i \,\varepsilon\, [D]$, where $i\colon K \to K_1$ is the identity. Also K is a D. R. of K_1 and $K_1 - K$ is simply connected. Therefore it follows from Lemma 10 that $i\colon K \to K_1$ is a simple equivalence. Since $k\colon K_1 \to K$ is a homotopy inverse of i if $k \,\varepsilon\, [D^{-1}]$, it follows that k is also a simple equivalence. Therefore $\phi\colon K \to DK$ is a simple equivalence, rel. L, if $\phi \,\varepsilon\, [D]$, where D is any elementary deformation, rel L. Therefore it follows from an inductive argument that $\phi\colon K \equiv K'$ (Σ), rel. L, if $K' = DK$, rel. L, and $\phi \,\varepsilon\, [D]$.

Conversely, let $\phi\colon K \equiv K'$ (Σ), rel. L. Then it follows from Theorem 11 and Lemma 12 that we may, without loss of generality, replace K' by K'', where $K'' \approx K'$, rel. L. Therefore we assume that $K \cap K' = L$ and also that $e^0 = e'^0 \,\varepsilon\, L$, where $e^0 \,\varepsilon\, K^0$, $e'^0 \,\varepsilon\, K'^0$ are the base points, thus excluding the case $L = 0$. Let P, with base point e^0, be the mapping cylinder of ϕ. Since $\phi\colon K \equiv K'$ the relations (14.1), (14.2) hold for every $n \geq 1$. Also K' is a D. R. of P. Therefore $i'_1\colon \pi_1(K') \approx \pi_1(P)$, where i'_1 is the injection. Therefore $C(K)$, $C(K')$ are sub-systems of $C(P)$, according to the convention (10.6). Let $j\colon C(L) \to C(P)$ be the chain mapping induced by the identical map $L \to P$. Then $A = jC(L) = C(K) \cap C(K')$. Let $h\colon C(K) \to C(K')$ be the chain mapping induced by ϕ. Then h is obviously rel. A, since ϕ is rel. L. It may be verified[38] that $C(P)$ is the mapping cylinder of h, as defined in the paragraph containing (8.4). Since, by hypothesis, h is a simple equivalence we have

$$C(P - K) \equiv 0\ (\Sigma). \tag{15.1}$$

according to (8.4).

---

[38] See Section 14 of CH(II), with $\gamma c_\lambda{}^n = 0$ if $c_\lambda{}^n$ corresponds to a cell in $L \subset K$ (not the L in CH(II)).

It follows from (14.1) that $\pi_r(P, K) = 0$ for every $r \geq 1$. Let

(15.2) $q = \text{Max}\{\dim(P - K), 3\} = \text{Max}\{\dim(K - L) + 1, \dim(K' - L), 3\}$.

Then it follows from Lemma 15 that there is a complex $Q = D_1 P$, rel. K, such that

$$Q = K \cup e_1^{q-1} \cup \cdots \cup e_s^{q-1} \cup e_1^q \cup \cdots \cup e_t^q. \tag{15.3}$$

By the first part of the Theorem, with K, K', L replaced by P, Q, K, we have $C(Q) \equiv C(P)$ (Σ), rel. $C(K)$. Therefore it follows from Theorem 4(a) and (15.1) that $C'' = C(Q - K) \equiv C(P - K) \equiv 0$ (Σ), whence $\partial''_q : C''_q \approx C''_{q-1}$ (Σ), by the corollary to Theorem 5. Therefore $s = t$. Let

$$\partial''_q m_i^q = \sum_{j=1}^{s} d_{ij} m_j^{q-1} \qquad (d_{ij} \varepsilon R),$$

where $(m_1^r, \cdots, m_s^r)$ is the basis of C''_r $(r = q - 1, q)$. Since $\tau(\partial''_q) = 0$ the matrix $\boldsymbol{d} = [d_{ij}]$ can be annihilated by an expansion

$$\boldsymbol{d} \to \begin{bmatrix} \boldsymbol{d} & 0 \\ 0 & \mathbf{1}_k \end{bmatrix},$$

followed by a sequence of elementary transformations of the form (2.12), followed by a contraction $\mathbf{1}_{s+k} \to \mathbf{1}_0$, where $\mathbf{1}_0$ is the empty matrix. Since $q - 1 \geq 2$ it follows from arguments on pp. 289, 290 of [1], with minor alterations,[39] that the transformation $\boldsymbol{d} \to \mathbf{1}_0$ can be "copied geometrically" by a formal deformation $Q \to K$, rel. K. Therefore $K = D_2 P$, rel. K, and the Theorem follows from the corollary to Lemma 11.

Let us describe r as the order of an elementary expansion $K \to K_1 = K \cup e^{r-1} \cup e^r$, and also of its inverse, $K_1 \to K$. Let $\phi : K \equiv K'$ (Σ), rel. L and let $K^p, K'^p \subset L$, for some $p \geq -1$. Then the following addendum to Theorem 13 is implicit in the proofs of Lemmas 11-15 and of Theorem 13.

ADDENDUM. *$K' = DK$, rel. L, and $\phi \varepsilon [D]$, where D is the resultant of elementary deformations, whose orders lie between $p + 2$ and $q + 1$ inclusive, where q is given by* (15.2).

This addendum has the following application. It follows from Theorems 11 and 13 that, by means of a formal deformation, we can reduce a given complex to one which has a given point, e^0, as its only 0-cell. It is sometimes convenient to restrict ourselves to a class of complexes, all of which have the same point, e^0, as their only 0-cell. Let K_0, K_r be two such complexes and let $K_r = DK_0$. Then it follows from Theorem 13 and its addendum, with $p = 0$, that $[D] = [D']$, where $D': K_0 \to K_r$ is the resultant of elementary deformation, $K_i \to K_{i+1}$, whose orders exceed 1. Therefore $K_j{}^0 = e^0$ for each $j = 0, \cdots, r$.

---

[39] Let $r(a_1, \cdots, a_k)$ mean the same as in Theorem 19 of [1], with $K^* = Q^{q-1}$ and $k = s$. Then $m_j^{q-1} \to a_j$ determines an isomorphism $C''_{q-1} \approx r(a_1, \cdots, a_k)$.

16. *n*-types. By a *cluster of n-spheres, attached to a space X, at a* point $x_0 \varepsilon X$, we shall mean a set of n-spheres, $\{S_i{}^n\}$, such that $X \cap S_i{}^n = x_0$ and $S_i{}^n - x_0$ does not meet $S_j{}^n$ if $i \neq j$. If X is a complex and $x_0 \varepsilon X^{n-1}$, then $X \cup \{S_i{}^n\}$ is the complex $X \cup \{e_i{}^n\}$, where $e_i{}^n = S_i{}^n - x_0$. At this stage we assume that, X being a finite complex, the number of n-spheres in a cluster attached to X is finite.

Let K^n, L^n be complexes of at most n dimensions $(n > 1)$ and let $\phi: K^n \to L^n$ be an $(n-1)$-homotopy equivalence, as defined in Section 2 of CH(I).

THEOREM 14. *There is a simple equivalence,*

$$\psi: K^n \cup \{S_{1i}{}^n\} \equiv L^n \cup \{S_{2j}{}^n\} \quad (\Sigma),$$

such that $\psi x = \phi x$ *if* $x \varepsilon K^{n-1}$, *where* $\{S_{1i}{}^n\}$, $\{S_{2j}{}^n\}$ *are clusters of n-spheres attached to* K^{n-1}, L^{n-1}.

Assuming that $K^n \cap L^n = 0$, let P be the mapping cylinder of ϕ. Then P^n is the union of K^n, L^n and the cells $e^r \times (0,1)$, where $e^r \varepsilon K^{n-1}$. We attach a cluster of n-spheres,

$$\{S_{2\rho}{}^n\} = S_{21}{}^n \cup \cdots \cup S_{2k}{}^n,$$

to a 0-cell $e'^0 \varepsilon L^0$, where k is to be determined later. Using Lemma 13, we transfer these over P^n to a 0-cell of K^n, so that they become a cluster, $\{S_\rho{}^n\}$, attached to K^{n-1}. Since $\phi: K^n \equiv_{n-1} L^n$ it follows that (14.1) and (14.2) are satisfied with $m = n - 1$. Therefore

(16.1) a) $\pi_r(P^n, K^n) = 0$ *if* $r = 1, \cdots, n-1$,
b) *the injection,* $\pi_{n-1}(K^n) \to \pi_{n-1}(P^n)$, *is an isomorphism (onto).*

These conditions are obviously satisfied by K^* and P^*, where

$$K^* = K^n \cup \{S_\rho\}, \qquad P^* = P^n \cup \{S_\rho{}^n\}.$$

Therefore it follows from (16.1a) and Lemma 15 that there is a formal deformation

$$D_1 : P^* \rightarrow Q = K^* \cup e_1^{n-1} \cup \cdots \cup e_a^{n-1} \cup e_1^n \cup \cdots \cup e_b^n, \quad \text{rel. } K^*.$$

Now let $k = a$. On considering the effect of a simple elementary deformation, rel. K^*, it follows inductively that (16.1) are also satisfied by K^* and Q. Let

$$g_\rho : (E_0^n, I^{n-1}) \rightarrow (Q, \bar{e}_\rho^{n-1}) \qquad (\rho = 1, \cdots, a)$$

mean the same as in (14.3). Since $g_\rho \mid \partial I^n = g_\rho \mid \partial E_0^n$ is homotopic in Q to a constant map, it follows from Lemma 13 that there is a formal deformation $Q \rightarrow Q'$, rel. K^n, which replaces each $S_\rho^n - e^0$ by a cell, $e_{b+\rho}^n$, with a characteristic map $g'_\rho : I^n \rightarrow \bar{e}_{b+\rho}^n$, such that $g'_\rho \mid \partial I^n = g_\rho \mid \partial I^n$. Therefore it follows, as in the proof of Lemma 15, that there is a formal deformation

$$D_2 : Q' \rightarrow Q'' = K^n \cup e'^n_1 \cup \cdots \cup e'^n_b, \quad \text{rel. } K^n.$$

Let $h : I^n \rightarrow \bar{e}'^n_i$ be a characteristic map for e'^n_i. Then it follows from (16.1b) that $h_i \mid \partial I^n$ is homotopic in K^n to a constant map. Therefore it follows from Lemma 13 that there is a formal deformation

$$D_3 : Q'' \rightarrow K^n \cup S_{11}^n \cup \cdots \cup S_{1b}^n, \quad \text{rel. } K^n,$$

where $\{S_{1i}^n\}$ is a cluster of n-spheres attached to K^{n-1}. On reversing these constructions we have

$$P^n \cup S_{21}^n \cup \cdots \cup S_{2k}^n = D(K^n \cup S_{11}^n \cup \cdots \cup S_{1b}^n), \quad \text{rel. } K^n.$$

Let $P_0^n \subset P^n$ be the mapping cylinder of $\phi \mid K^{n-1}$. Then P^n is the union of P_0^n and the n-cells in K^n. Let $f : I^n \rightarrow \bar{e}^n$ be a characteristic map for an n-cell $e^n \varepsilon K^n$. Then $f \mid \partial I^n$ is obviously homotopic in P_0^n to $\phi(f \mid \partial I^n) : \partial I^n \rightarrow L^{n-1}$. Since the latter can be extended to $\phi f : I^n \rightarrow L^n$ it follows that $f \mid \partial I^n$ is homotopic in P_0^n to a constant map. Therefore

$$P^*_0 = P_0^n \cup S_{21}^n \cup \cdots \cup S_{2l}^n = D'(P^n \cup S_{21}^n \cup \cdots \cup S_{2k}^n), \quad \text{rel. } P_0^n,$$

where $l \geq k$ and $S_{2,k+1}, \cdots, S_{2l}^n$ are n-spheres, attached to $e'^0 \varepsilon L^n$, which correspond to the n-cells in K^n. It follows from Lemma 11 that

$$L^* = L \cup S_{21}^n \cup \cdots \cup S_{2l}^n = D^* P_0^*,$$

where D^* is a contraction. Let $\psi^* : P_0^* \rightarrow L^*$ be given by $\psi^* \mid L^* = 1$,

$\psi^*(x, t) = \phi x$ $(x \varepsilon K^{n-1})$. Then $\psi^* \varepsilon [D^*]$ and the conditions of the theorem are satisfied by a map $\psi \varepsilon [D^*D'D]$. This completes the proof.

It follows from this theorem that any two complexes of the same n-type can be interchanged by a finite sequence of elementary deformations and transformations of the form

$$(16.2) \qquad K \rightarrow L = K - e^r, \qquad L \rightarrow K = L \cup e^r \qquad (r > n),$$

where e^r is a principal cell of K. For the transformations $K \rightarrow K^n \rightarrow K^n \cup E^{n+1} \rightarrow K^n \cup S^n$ are the resultants of such sequences, where $E^{n+1} = e^0 \cup e^n \cup e^{n+1}$, $S^n = \partial E^{n+1} = e^0 \cup e^n$, and $K^n \cap E^{n+1} = e^0 \varepsilon K^0$. Conversely a formal deformation preserves the homotopy type, and hence the n-type of a complex. Also $K^n = L^n$, whence K, L are of the same n-type, if they are related by (16.2). Thus the n-type may be defined in terms of formal deformations and elementary transformations of the form (16.2).

It follows from Lemma 2 in Section **9** of CH(I) and from Lemma 13 and Theorem 12 above that, if K is any complex, there is a simplicial complex $K^* = DK$. Moreover, Sections **14-16** may be interpreted as referring to formal deformations of the kind considered in [3]. Therefore the class of simplicial complexes, which, when treated as cell-complexes, are of the same simple homotopy type, or n-type, as a given one, K, is the same as the "nucleus," or "n-group," of K, as defined in [1].

17. Homotopy systems.[40] We proceed to the simple equivalence theory of homotopy systems. In this section we confine ourselves to systems, ρ, such that $\dim \rho < \infty$ and each group ρ_n has a finite basis.

We modify the definition of a homotopy system, ρ, by associating a class of *preferred bases* with each ρ_n. Let $(a_1, \cdots, a_p)$ be a preferred basis for ρ_n. Then $(a'_1, \cdots, a'_p)$ shall be a preferred basis for ρ_n if, and only if, $a'_i = a_{j_i}^{\pm 1}$, in case $n = 1$, or $a'_i = \pm x_i a_{j_i}$ if $n > 1$, where $x_i \varepsilon \rho_1$ and $j_1, \cdots, j_p$ is a permutation of $1, \cdots, p$. We shall only admit that $f : \rho \approx \rho'$ if f carries a preferred basis for each ρ_n into a preferred basis for ρ'_n. The preferred bases for $\rho(K)$, where K is a complex, shall be the natural bases, as defined in Section **5** of CH(II). If K^0 is a single 0-cell, then the natural bases for $\rho_1(K)$ are uniquely defined. In general they depend on the choice of a tree $T \subset K^1$, which contains K^0. In this case $\rho(K)$ is the homotopy system of the pair, (K, T). However we shall continue to write it as $\rho(K)$. A complex, K, will be called a (geometrical) realization of a given system, ρ, if, and only if, $\rho \approx \rho(K)$, subject to our condition concerning preferred bases. The process of realizing ρ by K will consist of defining a particular $f: \rho \approx \rho(K)$. Having (implicitly) done this, we shall use ρ and $a \varepsilon \rho$ to denote $\rho(K)$ and fa. By a *basis* for ρ_n we shall always mean a preferred basis.

---

[40] The main purpose of this section is to prove Theorem 17, which was announced in Section 7 of CH(II).

Let C and $h: \rho \rightarrow C$ be defined as in Section 8 of CH(II). Then C shall be a chain system of the kind introduced in Section 2 above, R being the group ring of $\bar{\rho}_1 = \rho_1/d\rho_2$. We insist that, if $(a_1, \cdot \cdot \cdot, a_p)$ is a (preferred) basis for ρ_n and if $(m'_1, \cdot \cdot \cdot, m'_p)$ is the basis of C_n, then $ha_i = \pm \bar{x}_i m'_{j_i}$, where $\bar{x}_i \varepsilon \bar{\rho}_1$ and $\bar{x}_i = 1$ if $n = 1$.

We are going to define a sub-system, ρ', of a homotopy system ρ. This is not quite so simple as in the case of chain systems, for the following reason. Let $\rho'_1 \subset \rho_1$ be the sub-group generated by part of a basis for ρ_1. Let $\rho'_2 \subset \rho_2$ be the sub-group generated by ρ'_1, operating on a set of elements, $(a'_1, \cdot \cdot \cdot, a'_k)$, in a basis for ρ_2. Let $d\rho'_2 \subset \rho'_1$ and let $d': \rho'_2 \rightarrow \rho'_1$ be the homomorphism induced by $d: \rho_2 \rightarrow \rho_1$. Then ρ'_2 is not necessarily a free crossed (ρ'_1, d')-module. For example, let ρ_1 have a single free generator, x, and ρ_2 a pair of basis elements, a, b, such that $da = x$, $db = 1$. Let $\rho'_1 = \rho_1$ and let ρ'_2 be generated by ρ_1, operating on b. Since $db = 1$ we have $a + b = b + a$, whence $xb - b = a + b - a - b = 0$. Therefore ρ'_2 is not a free ρ_1-module.

Let $\bar{\rho}'_1 = \rho'_1/d\bar{\rho}'_2$ and let $i_*: \bar{\rho}'_1 \rightarrow \bar{\rho}_1$ be the homorphism induced by the identical map $i_1: \rho'_1 \rightarrow \rho_1$.

LEMMA 16. *Let* $i_*^{-1}(1) = 1$. *Then* ρ'_2 *is a free crossed* (ρ'_1, d')*-module, having* $(a'_1, \cdot \cdot \cdot, a'_k)$ *as a basis.*

Let ρ''_2 be the free crossed (ρ'_1, d'')-module, which is defined in terms of the symbolic generators (x', α_i) and the map $\alpha_i \rightarrow da'_i$ $(i = 1, \cdot \cdot \cdot, k$; $x' \varepsilon \rho'_1)$. Obviously $d''\rho''_2 = d\rho'_2$. Let $a''_i \varepsilon \rho''_2$ be the basis element which corresponds to the generator $(1, \alpha_i)$. It follows from Lemma 2 in Section 2 of CH(II) that an operator homomorphism, $i_2: \rho''_2 \rightarrow \rho_2$, associated with $i_1: \rho'_1 \rightarrow \rho_1$, is defined by $i_2 a''_i = a'_i$. Obviously $i_2\rho''_2 = \rho'_2$ and the lemma will follow when we have proved that $i_2^{-1}(0) = 0$.

Let $C_2 = h\rho_2$, $C''_2 = h''\rho''_2$ be ρ_2, ρ''_2 made Abelian and let $j: C''_2 \rightarrow C_2$ be the homomorphism induced by i_2. Since $a'_i = i_2 a''_i$ and $jh'' = hi_2$ it follows that $(jh''a''_1, \cdot \cdot \cdot, jh''a''_k)$ is part of a basis for C_2. Since $d''\rho''_2 = d\rho'_2$ and $i_*^{-1}(1) = 1$ it follows that $j^{-1}(0) = 0$. Let $a'' \varepsilon i_2^{-1}(0)$. Then

$i_1 d'' a'' = d i_2 a'' = 1$, $j h'' a'' = h i_2 a'' = 0$. Therefore $d'' a'' = 1$, $h'' a'' = 0$ and it follows from Lemma 1 in CH(II) that $a'' = 0$. Therefore $i_2^{-1}(0) = 0$ and the lemma is proved.

Let ρ'_1, ρ'_2 satisfy the conditions of Lemma 16. Let $\rho'_p \subset \rho_p$ $(p = 3, 4, \cdots)$ be the sub-group which is generated by $\bar{\rho}'_1$, operating on part of a basis for ρ_p, and let $d\rho'_p \subset \rho'_{p-1}$. Let $d' : \rho'_p \to \rho'_{p-1}$ be the homomorphism induced by $d : \rho_p \to \rho_{p-1}$. Then $\rho' = \{\rho'_p\}$, with d' as boundary operator, is a homotopy system, which we describe as a *sub-system* of ρ, on the understanding that a (preferred) basis for ρ'_n $(n \geqq 1)$ is part of a basis for ρ_n.

Let ρ be a given homotopy system, let $Z_n(\rho) = d_n^{-1}(0)$ and let[41]

$$G_1(\rho) = \bar{\rho}_1, \qquad G_n(\rho) = Z_n(\rho) - d_{n+1}\rho_{n+1} \qquad (n > 1).$$

A homomorphism, $f : \rho \to \rho'$, obviously induces a family of homomorphisms $f_* : G_n(\rho) \to G_n(\rho')$ $(n = 1, 2, \cdots)$. It may be verified in the same way as in ordinary homology theory that $f_* : G_n(\rho) \approx G_n(\rho')$ if $f : \rho \equiv \rho'$. The converse is proved below.

Let $\rho' \subset \rho$ be a sub-system and let $Z_n(\rho, \rho') = d_n^{-1}\rho'_{n-1}$ $(n > 1)$. Let $a \,\varepsilon\, Z_2(\rho, \rho')$, $a' \,\varepsilon\, \rho'_2$. Then $a + a' - a = (da)a' \,\varepsilon\, \rho'_2$, since $da \,\varepsilon\, \rho'_1$. Therefore ρ'_2 is an invariant sub-group of $Z_2(\rho, \rho')$. So therefore is the direct sum $\rho'_2 + d\rho_3$, since $d\rho_3 \subset Z_2(\rho)$, which is in the centre of ρ_2. Let

$$G_n(\rho, \rho') = Z_n(\rho, \rho') - (\rho'_n + d\rho_{n+1}) \qquad (n > 1).$$

Let

$$(17.1) \qquad G_n(\rho') \xrightarrow[i_*]{} G_n(\rho) \xrightarrow[j_*]{} G_n(\rho, \rho') \xrightarrow[d_*]{} G_{n-1}(\rho') \xrightarrow[i_*]{} \cdots \xrightarrow[i_*]{} G_1(\rho)$$

be the homomorphisms, which are induced by $i : \rho' \to \rho$, the identical map $Z_n(\rho) \to Z_n(\rho, \rho')$ and by $d \mid Z_n(\rho, \rho')$. Then it may be verified, as in ordinary homology theory, that the sequence (17.1) is exact.

Let $f : \rho \to \rho'$ be a homomorphism of ρ into a system ρ', with boundary operator d'. Let $f_* : \bar{\rho}_1 \approx \bar{\rho}'_1$. We proceed to define a system, ρ^*, which we shall call the *mapping cylinder* of f. We realize the systems $\rho^2 = (\rho_1, \rho_2)$ and ρ'^2 by complexes $K = K^2$ and $K' = K'^2$, such that $K^0 = K'^0 = e^0 = K \cap K'$. By Theorem 4 in CH(II), $f : \rho^2 \to \rho'^2$ can be realized by a map $\phi : K \to K'$. Let P be the mapping cylinder of ϕ, with $e^0 \times I$ shrunk into the point e^0. Then $P^0 = e^0$. We define $\rho^*_n = \rho_n(P) = \rho_n(P^2)$ $(n = 1, 2)$.

Since ϕ induces $f_* : \bar{\rho}_1 \approx \bar{\rho}'$ and since K' is a D. R. of P, it follows that

---

[41] If $\rho = \rho(K)$ then $G_1(\rho) \approx \pi_1(K)$, $G_2(\rho) \approx \pi_2(K)$, $G_n(\rho) \approx H_n(K)$ if $n > 2$.

$i_*: \bar{\rho}_1 \approx \bar{\rho}^*_1$, $i'_*: \bar{\rho}'_1 \approx \bar{\rho}^*_1$, where i_*, i'_* are the injections. Therefore it follows from the proof of Lemma 16 that the injections

$$(17.2) \qquad i: \rho^2 \to \rho(P^2), \qquad i': \rho'^2 \to \rho(P^2)$$

are isomorphisms (into).

Let $\delta: \rho^2 \to \rho(P)$ be the deformation operator determined by the homotopy $\delta_t: K \to P$, which is given by $\delta_t p = (p, t)$ $(p \in K)$. Then

$$(17.3) \qquad d^*_n \delta_n = f_{n-1} - 1 - \delta_{n-1} d_n \qquad (\delta_1 d_1 = 0),$$

where $n = 2, 3$ and ρ^*_1 is written additively. Let $n \geqq 3$ and let $\delta_n \rho_{n-1}$ be a free $\bar{\rho}^*_1$-module, which is the image of ρ_{n-1} in an operator homomorphism, δ_n, whose kernel is the commutator sub-group of ρ_{n-1}. Thus $\delta_n: \rho_{n-1} \approx \delta_n \rho_{n-1}$ if $n > 3$. We take $\delta_3 \delta_2 \subset \rho_3(P)$ and δ_3 shall mean the same as before. Let ρ^*_n be the direct sum $\rho^*_n = \rho'_n + \rho_n + \delta_n \rho_{n-1}$ $(n \geqq 3)$. We imbed ρ_n, ρ'_n in ρ^*_n by means of $i: \rho_n \to \rho^*_n$, $i': \rho'_n \to \rho^*_n$, where i, i' mean the same as in (17.2) if $n = 1, 2$, and $ia = (0, a, 0)$, $i'a' = (a', 0, 0)$ if $n > 2$. We define $d^*_n: \rho^*_n \to \rho^*_{n-1}$ by $d^*_n a = d_n a$, $d^*_n a' = d'_n a'$ and by (17.3), with $n \geqq 2$. If $\{a_i^n\}$ and $\{a'_j{}^n\}$ are bases for ρ_n and ρ'_n, then the union of $\{a_i^n\}$, $\{a'_j{}^n\}$ and $\{\delta_n a_k^{n-1}\}$ shall be a (preferred) basis for ρ^*_n. It follows from an argument in Section 8 above that $d^*_n d^*_{n+1} = 0$ if $n \geqq 3$. Also $d^* d^* = 0$ in $\rho(P)$. Therefore $d^*_n d^*_{n+1} = 0$ for every $n > 1$. Clearly d^*_n is an operator homomorphism and it follows that $\rho^* = \{\rho^*_n\}$, with $d^* = \{d^*_n\}$ as boundary operator, is a homotopy system. We call it the *mapping cylinder* of $f: \rho \to \rho'$.

Let $i': \rho' \to \rho^*$ be the identical map and let $k': \rho^* \to \rho'$ be given by

$$k'a = fa, \qquad k'a' = a', \qquad k'\delta a = 0 \qquad (a \in \rho, a' \in \rho').$$

Then $k'i' = 1$ and it is easily verified that $d'k' = k'd^*$ and that $i'k' - 1 = d^*\delta^* + \delta^* d^*$, where $\delta^* a = \delta a$, $\delta^* \rho' = \delta^* \delta \rho = 0$. Therefore $k': \rho^* \equiv \rho'$ and $k'_*: G_n(\rho^*) \approx G_n(\rho')$. Clearly $f = k'i$, where $i: \rho \to \rho^*$ is the identical map, whence $f_* = k'_* i_*$. Therefore, if each f_* is an isomorphism (onto), so is i_*. In this case it follows from the exactness of (17.1) that

$$(17.4) \qquad G_n(\rho^*, \rho) = 0 \qquad (n \geqq 1),$$

where $G_1(\rho^*, \rho) = 0$ means that $\rho^*_1 = i_* \rho_1$.

Let $r \geqq 2$ and let $(a_0^n, a_1^n, \cdots, a_{p_n}^n)$ be a preferred basis for ρ_n $(n = r - 1, r)$. If $r = 2$ let $\rho'_1 \subset \rho_1$ be the sub-group generated by $(a_1^1, \cdots, a_{p_1}^1)$ and if $r > 2$ let $\rho'_1 = \rho_1$. If $n > 1$ $(n = r - 1, r)$ let

$\rho'_n \subset \rho_n$ be the sub-group which is generated by ρ'_1 operating on $(a_1^n, \cdots, a_{p_n})$. In any case let

$$da_0^r = a_0^{r-1} - a'_0, \qquad da_i^r \in \rho_{r-1} \qquad\qquad (i = 1, \cdots, p_r),$$

where $a'_0 \in \rho'_{r-1}$ and ρ_1 is written additively if $r = 2$. Then $d\rho'_r \subset \rho'_{r-1}$ and the conditions of Lemma 16 are satisfied[42] by ρ'_1, ρ'_2. Therefore $\rho' = \{\rho'_n\}$, with $\rho'_n = \rho_n$ if $n \neq r - 1$ or r, is a sub-system of ρ. Let $i : \rho' \rightarrow \rho$ be the identical map and let $k_n : \rho_n \rightarrow \rho'_n$ the operator homomorphism, which is given by $k_n \mid \rho'_n = 1$ and $k_r a_0^r = 0$, $k_{r-1} a_0^{r-1} = a'_0$. Then it is easily verified that $kd = d'k$, whence $k : \rho \rightarrow \rho'$ is a homomorphism. We have $ki = 1$ and $ik - 1 = d\xi + \xi d$, where $\xi : \rho \rightarrow \rho$ is the deformation operator given by

$$\xi\rho' = 0, \qquad \xi a_0^r = 0, \qquad \xi a_0^{r-1} = - a_0^r.$$

Therefore $i : \rho' \equiv \rho$ and $k : \rho \equiv \rho'$. Notice that, if $f : \rho \rightarrow \rho'$ is any homomorphism such that $fi \simeq 1$, then $f \simeq fik \simeq k$.

We shall describe a homomorphism $f : \rho^0 \rightarrow \rho^1$ as an *elementary equivalence* if, and only if, ρ^0, ρ^1 are related to each other in the same way as ρ, ρ' in the preceding paragraph, and $f \cong i$ or $f \cong k$, according as $\rho^0 \subset \rho^1$ or $\rho^1 \subset \rho^0$. We shall describe a homomorphism $f : \rho \rightarrow \rho^*$ as a *simple equivalence*, $f : \rho \equiv \rho_0\ (\Sigma)$, if, and only if, it is the resultant of a finite sequence of isomorphisms and elementary equivalences.

Let C, C' be chain systems associated with given homotopy systems ρ, ρ^1. Let $g : C \rightarrow C'$ be the chain mapping induced by a homomorphism $f : \rho \rightarrow \rho'$.

THEOREM 15. *$f : \rho \equiv \rho'\ (\Sigma)$ if, and only if, $g : C \equiv C'\ (\Sigma)$.*

This follows from the lemmas in Section **14** and the proof of Theorem 13, restated in terms of homotopy ssytems.

Let ρ be a homotopy system and σ a free $\bar{\rho}_1$-module, with a finite basis $(b_1, \cdots, b_q)$. Let $\rho_n^0 = \rho_n + \sigma$, $\rho_p^0 = \rho_p$ $(p \neq n)$, for a given value of $n \geqq 2$. Let $d^0 : \rho_r^0 \rightarrow \rho_{r-1}^0$ be defined by $d^0 \mid \rho = d$, $d^0\sigma = 0$. Then[43] $\rho^0 = \{\rho_r^0\}$, with d^0 as boundary operator, is a homotopy system. We say that $\rho \rightarrow \rho^0$ is the result of *attaching a cluster* of n-cycles to ρ. If $\{a_i\}$ is a preferred basis for ρ_n, then $\{a_i, b_j\}$ shall be a preferred basis for ρ_n^0 and the preferred bases for ρ_p $(p \neq n)$ shall be the same in ρ^0 as in ρ.

---

[42] The homomorphism $k_1 : \rho_1 \rightarrow \rho'_1$, defined below, induces $i_*^{-1} : \rho_1 \approx \rho'_1$.

[43] If $n = 2$ then ρ_n^0 is a free crossed module since $d^0\sigma = 1$.

Let $\dim \rho$, $\dim \rho' \leq n$ $(n \geq 2)$ and let $f: \rho \rightarrow \rho'$ be a homomorphism such that

$$(17.5) \qquad f_*: G_r(\rho) \approx G_r(\rho') \qquad (r = 1, \cdots, n-1).$$

Then we have the following generalization of Tietze's theorem.

THEOREM 16. *There is a simple equivalence,* $f^0: \rho^0 \equiv \rho'^0$ (Σ), *such that* $f^0 a = fa$ *if* $a \varepsilon \rho_r$ $(r < n)$, *where* ρ^0, ρ'^0 *are formed by attaching clusters of* n*-cycles to* ρ, ρ'.

Since $f_*: \bar{\rho}_1 \approx \bar{\rho}'_1$ we can construct the mapping cylinder, ρ^*, of f. Then the theorem follows from the proof of Theorem 14, translated into algebraic terms.

The following corollary may be deduced from Theorem 16, or proved directly with the help of (17.4).

COROLLARY. *If* $\dim \rho$, $\dim \rho' \leq n-1$, *then* (17.5) *implies* $f: \rho \equiv \rho'$.

THEOREM 17. *If* $f: \rho(K) \equiv \rho'$, *where* K *is a complex, then* ρ' *can be realized by a complex,* K', *and in such a way*[44] *that* f *has a realization* $\phi: K \equiv K'$.

This follows from Theorem 16 and an argument which is essentially the same as the proof of Theorem 9 on p. 1228 of [3].

18. Infinite complexes. Let K_1 be a CW-complex, as defined in CH(I), which may be infinite. Let $K_0 \subset K_1$ be a sub-complex such that $K_1 = K_0 \cup \bigcup_{\alpha} (e_\alpha^{n-1} \cup e_\alpha^n)$, where $\{e_\alpha^{n-1}, e_\alpha^n\}$ is an indexed aggregate of cells such that $e_\alpha^{n-1} \cup e_\alpha^n$ is an open subset of K_1 and $K_0 \rightarrow K_0 \cup e_\alpha^{n-1} \cup e_\alpha^n$ is an elementary expansion, for each α. Then $K_0 \rightarrow K_1$ will be called a *composite expansion* and $K_1 \rightarrow K_0$ a *composite contraction.* It follows from the argument used in the finite case, and (I), in Section **5** of CH(I), that K_0 is a D.R. of K_1. By a *formal deformation,* $D: K \rightarrow L$, we shall mean the resultant of a finite sequence of composite expansions and contractions. We restrict ourselves to complexes of finite dimensionality. Then the proofs of the lemmas in Section **14** and of Theorem 14 apply to infinite complexes, after a few trivial alterations.

We also admit homotopy systems of finite dimensionality, in which the groups may have infinite bases. We define a simple equivalence, $f: \rho \equiv \rho'$, where ρ, ρ' are two such systems, by analogy with a formal deformation $D: K \rightarrow L$. Then Theorems 16, 17 can be extended without difficulty to systems in which the bases may be infinite.

---

[44] In general ρ' can also be realized by a complex, K', in such a way that f has no realization $K \rightarrow K'$.

It remains to be seen whether or not the purely algebraic theory developed in Section 2-9 can be extended to systems of modules with infinite bases, in such a way as to yield a generalization of Theorem 13 to infinite complexes.

MAGDALEN COLLEGE, OXFORD.

REFERENCES.

1. J. H. C. Whitehead, *Proceedings of the London Mathematical Society*, vol. 45 (1939), pp. 243-327.
2. ———, *Annals of Mathematics*, vol. 41 (1940), pp. 809-824.
3. ———, *ibid.*, vol. 42 (1941), pp. 1197-1239.
4. ———, *Bulletin of the American Mathematical Society*, vol. 54 (1948), pp. 1125-32.
5. ———, *ibid.*, vol. 54 (1948), pp. 1133-45.
6. N. Jacobson, *The theory of rings*, New York (1943).
7. K. Reidemeister, *Einführung in die kombinatorische Topologie*, Brunswick (1932).
8. ———, *Journal für die reine und angewandte Mathematik*, vol. 173 (1935), pp. 164-173.
9. W. Franz, *ibid.*, vol. 176 (1937), pp. 113-134.
10. G. de Rham, *Commentarii Mathematici Helvetici*, vol. 12 (1940), pp. 191-211.
11. G. W. Higman, *Proceedings of the London Mathematical Society*, vol. 46 (1940), pp. 231-248.
12. M. H. A. Newman, *K. Akademie van Wetenschappen, Amsterdam*, vol. 29 (1926), pp. 611-626; 627-641.
13. J. W. Alexander, *Annals of Mathematics*, vol. 31 (1930), pp. 292-320.
14. S. Lefschetz and J. H. C. Whitehead, *Transactions of the American Mathematical Society*, vol. 35 (1933), pp. 510-517.
15. J. L. Kelley and Everett Pitcher, *Annals of Mathematics*, vol. 48 (1947), pp. 682-709.
16. R. H. Fox, *ibid.*, vol. 44 (1943), pp. 40-50.

LIE ALGEBRAS AND DIFFERENTIATIONS IN RINGS OF POWER SERIES.*

By G. Hochschild.

Introduction. It was proved by Ado in 1934, [1], that every Lie algebra over a field of characteristic zero can be faithfully represented by linear transformations in a finite dimensional vector space. The details of this proof, which was published in Russian, are not widely known. Ado's proof is believed to be incomplete in one point and has the further disadvantage that it is rather elaborate and does not lead to a straightforward construction of a faithful representation for a given Lie algebra.

In 1938, E. Cartan published an entirely different proof of Ado's theorem, [3], for the case where the basic field is the field of the complex numbers. In fact, Cartan proves directly the stronger result that there exists a Lie group of linear transformations whose Lie algebra is isomorphic with the given Lie algebra, and that this group can be taken to be simply connected if the Lie algebra is solvable. Cartan's proof makes use of analysis and the theory of Lie groups of transformations, but the theorem for arbitrary base fields of characteristic zero could be obtained rather easily from Cartan's theorem.

Very recently, Harish-Chandra, [7], has given a purely algebraic proof of Ado's theorem by perfecting a method which previously had been successful only in the nilpotent case (G. Birkhoff, [2]) and for 'restricted' Lie algebras of characteristic p (Jacobson, [8]).

What we shall do here is to make the analytical tools used by Cartan available to algebra by a systematic use of the theory of differential forms in rings of formal power series. More specifically, we shall apply the algebraic version of what is known as the differential calculus of Cartan[1] to show that the elements of a given Lie algebra can be represented as differentiations in a ring of power series which map a finite dimensional subspace into itself and thus yield a faithful linear representation.

By a slight modification of Cartan's procedure this program can be carried out directly for the solvable case, i. e., the preliminary study of

---

* Received May 2, 1948.

[1] This will be found, for instance in [4], ch. V, and in [6]. The definitions we shall give in 2 differ from those in [4] and [6] only in some inessential respects.

nilpotent Lie algebras can be short circuited. The main new difficulty arises in extending the construction to the general case. This necessitates a close study of a certain system of differential equations which is "solvable" in the sense of Lie's classical theory.

In **1** we prove a few elementary results concerning Lie algebras. These are required in the passage from a representation of the radical of a Lie algebra to a representation of the whole Lie algebra. In **2** we give an outline of the theory of differential forms which is fundamental in all the later constructions. The representation of solvable Lie algebras is dealt with in **3**, and the extension to the general case is carried out in **4**, **5**.

1. Lie algebras. We shall require a few auxiliary results concerning Lie algebras. The first of these is an easy generalization of Levi's theorem:[2]

THEOREM 1.1. *Let P be a semisimple Lie algebra over a field K of characteristic 0. Suppose that H is another Lie algebra over K and that π is a homomorphism of H onto P. Then there exists an isomorphism τ of P into H such that $\pi\tau$ is the identity mapping on P.*

Proof. Let Q denote the kernel of π, R the maximal solvable ideal of H. The image under π of the sum (Q,R) is an ideal in P. Since it is isomorphic with $(Q,R)/Q$, or with $R/R \cap Q$, it is solvable. Since P is semisimple, this implies that $(Q,R)/Q = (0)$, i. e., $R \subseteq Q$.

On the other hand, H/R is semisimple and Q/R is an ideal in H/R. Hence H/R is the direct sum of Q/R and a complementary ideal which we may write S/R, where S is an ideal in H and contains R. Evidently, π maps S onto P, and the kernel of the restriction of π to S is $S \cap Q = R$. Now S/R, as a non-zero ideal of the semisimple algebra H/R, is semisimple, whence R is the maximal solvable ideal of S.

By Levi's theorem, we have a linear decomposition $S = T + R$, where T is a subalgebra which is isomorphic with S/R and therefore with P. It is now clear that there exists an isomorphism τ of P onto T which satisfies the condition of our theorem.

The next theorem is a refinement of Levi's theorem:

THEOREM 1.2. *Let L be a Lie algebra over K, R its maximal solvable ideal. Then L is the direct sum of two ideals T and H, such that*

---

[2] Levi's theorem states that a Lie algebra L over a field of characteristic O can be decomposed into a linearly direct sum $S + H$, where H is the maximal solvable ideal of L and S is a semisimple subalgebra. An elegant proof is given by J. H. C. Whitehead in [9].

(1) *T is semisimple or* (0);

(2) *H contains R as its maximal solvable ideal, and if $H = P + R$ is a Levi decomposition, then no non-zero element of P effects an inner derivation in R, i. e., if $0 \neq p \varepsilon P$ there exists no r_p in R such that $r \circ p = r \circ r_p$, for every $r \varepsilon R$.*

Proof. Applying Levi's theorem we obtain a linear decomposition $L = S + R$, where S is either (0), in which case there is nothing to prove, or a semisimple subalgebra of L.

Now denote by T the set of all $s \varepsilon S$ for which there is an $r_s \varepsilon R$ with $r \circ s = r \circ r_s$, for all $r \varepsilon R$. It is easy to verify that T is an ideal in S. Hence T is either (0) or semisimple.

Let Z denote the center of R. Then, for any $t \varepsilon T$, the elements $r_t \varepsilon R$ for which $r \circ t = r \circ r_t$, for all $r \varepsilon R$, make up exactly one coset $\bar{t}$ mod Z. The mapping $t \rightarrow \bar{t}$ is easily seen to be a homomorphism of T into R/Z. Since R/Z is solvable, we must have $\bar{T} = (0)$, for otherwise it would be a semisimple subalgebra of R/Z. This means that $R \circ T = (0)$.

Since S is semisimple, it is the direct sum of T and a complementary ideal P. Hence T is a direct summand in L, and $H = P + R$ is a complementary ideal in L which satisfies condition (2) of our theorem. In fact, if one Levi decomposition of H has the property described in (2), then every Levi decomposition will have this property.

We shall later have to make reference to the maximal nilpotent ideal of a Lie algebra. In the remainder of this section we give a few facts concerning this concept.

Let us recall that a Lie algebra N is said to be nilpotent if, with $N_0 = N$, and $N_{i+1} = N_i \circ N$, there is an n such that $N_n = (0)$. As we shall see below, the sum of all nilpotent ideals of a Lie algebra is nilpotent, and it is called the maximal nilpotent ideal.

We shall take the following fundamental theorem for granted ([5], th. 3).

THEOREM 1. 3. (Lie) *Let L be a solvable Lie algebra. Then every simple representation space for L is annihilated by the derived algebra $L \circ L$.*

Let M be an arbitrary representation space for the Lie algebra L. A subset S of L is said to be nilpotent on M if there exists an integer n such that, for every $m \varepsilon M$ and every set $s_1, \cdots, s_n$ of elements of S, we have $s_1 \cdots s_n \cdot m = 0$, or—as we shall indicate more briefly—if $S^n \cdot M = (0)$. It follows almost immediately from theorem 1. 3 that if L is solvable then

$L \circ L$ is nilpotent on every representation space of L. In fact, we merely have to consider a composition series for the L-module M and note that $L \circ L$ maps each term of this series into the next term, because the quotients of successive terms are simple L-modules.

LEMMA 1.1. *Let M be a representation space for the Lie algebra L, S a subspace of L which is nilpotent on M. Let x be an element of L which is nilpotent on M and such that $x \circ S \subseteq S$. Then the subspace (S, x) of L which is spanned by S and x is nilpotent on M.*

Proof. There are indices p and q such that $S^p \cdot M = (0)$ and $x^q \cdot M = (0)$. Since $s \cdot (x \cdot m) = x \cdot (s \cdot m) + (x \circ s) \cdot m$, it is easy to see that every element of $(S, x)^n \cdot M$ in which altogether t operators from S occur can be written as a sum of elements each of which belongs to an $x^r \cdot (S^t \cdot M)$, with some index r. Hence the elements in which $t \geqq p$ are zero. Thus, the total number of operators from S in a non-zero element of $(S, x)^n \cdot M$ must be less than p. But then, if the total number of x's is greater than $p(q-1)$, such an element must involve x^q in at least one place, and hence is zero. Hence we must have $(S, x)^n \cdot M = (0)$ as soon as $n \geqq p(q-1) + p = pq$.

We may regard L as a representation space for any subalgebra of L, such that the transform by an element x of the element $z \varepsilon L$ is given by $x \cdot z = D_x(z) = z \circ x$. With this understanding, we have:

LEMMA 1.2. *Let R be the maximal solvable ideal of the Lie algebra L. Then $R \circ L$ is nilpotent on L.*

Proof. If x is an arbitrary element of L then (R, x) is a solvable subalgebra of L, because $(R, x) \circ (R, x) = (R \circ R, R \circ x) \subseteq R$. Hence, by the above, $(R \circ R, R \circ x)$ is nilpotent on L. Since, for every $y \varepsilon L$, $(R \circ y) \circ (R \circ R, R \circ x) \subseteq R \circ R$, it follows by repeated applications of Lemma 1.1 that $R \circ L$ is nilpotent on L.

THEOREM 1.4. *Let L be a Lie algebra, R its maximal solvable ideal. Let N be the set of all elements of R which are nilpotent on L. Then N is a nilpotent ideal of L and coincides with the sum of all nilpotent ideals of L.*

Proof. Clearly, N must contain every nilpotent ideal of L. In particular, $N \supseteq R \circ L$. If $x \varepsilon N$ it follows from Lemma 1.1 that the subspace $(x, R \circ L)$ of L is contained in N. By repeating this argument a finite number of times, noting that $y \circ (x, R \circ L) \subseteq R \circ L$, etc., we finally conclude that N is a subspace of R. It is then clear that N is actually a nilpotent ideal of L. The remaining assertion of our theorem is a trivial consequence.

COROLLARY. *The maximal nilpotent ideal of L coincides with the maximal nilpotent ideal of R.*

Proof. An element of R is nilpotent on R if and only if it is nilpotent on L.

THEOREM 1.5. *Let L, R, N be as above, and let D be a derivation in R. Then $D(R) \subseteq N$.*

Proof. We construct a Lie algebra R^* as follows: The vector space of R^* is the direct sum of R and the basic field K, i. e., the elements of R^* are pairs (x, k), where $x \varepsilon R$ and $k \varepsilon K$. We define $(x_1, k_1) \circ (x_2, k_2) = (x_1 \circ x_2 + k_2 D(x_1) - k_1 D(x_2), 0)$. Then R^* is easily seen to be a solvable Lie algebra which contains $R = (R, 0)$ as an ideal. Hence we may regard R as a representation space for R^*. Since R^* is solvable, we conclude that $R^* \circ R^* = (R \circ R, D(R))$ (subspace of R) is nilpotent on R, and hence on L. Hence $D(R) \subseteq N$.

2. Differentiations and differential forms.

Let K be a field of characteristic 0. We form the ring $K\langle x_1, \cdots, x_n\rangle$ of integral (i. e., with exponents $\geqq 0$) formal power series in n variables $x_1, \cdots, x_n$ over K.

Definition 2.1. A differentiation in $K\langle x_1, \cdots, x_n\rangle$ is a mapping D of $K\langle x_1, \cdots, x_n\rangle$ into itself such that

(1) $D\{a\} = 0$, for every $a \varepsilon K$;

(2) For any two power series p, q, we have

$$D\{p + q\} = D\{p\} + D\{q\}, \text{ and } D\{pq\} = D\{p\}q + pD\{q\}.$$

Evidently, every differentiation is a K-linear transformation. Denote by $D_1, \cdots, D_n$ the partial differentiations (in the ordinary sense) with respect to $x_1, \cdots, x_n$, respectively. Then the D_i are evidently differentiations in the above sense. In fact, we have the following theorem:

THEOREM 2.1. *The set of differentiations in $K\langle x_1, \cdots, x_n\rangle$ coincides with the set of all mappings of the form $\sum_{i=1}^{n} p_i D_i$, where the p_i are arbitrary elements of $K\langle x_1, \cdots, x_n\rangle$. Hence the differentiations constitute a free $K\langle x_1, \cdots, x_n\rangle$-module of rank n.*

Proof. Let N denote the ideal of all non-units in $K\langle x_1, \cdots, x_n\rangle$.

Then N coincides with the ideal generated by $x_1, \cdots, x_n$, and $\bigcap_{e=1}^{\infty} N^e = (0)$. Clearly, if D is any differentiation, and $e > 1$, then $D(N^e) \subseteq N^{e-1}$.

Given a differentiation D, consider the differentiation

$$D' = D - \sum_{i=1}^{n} D\{x_i\} D_i.$$

It is clear that D' maps every polynomial (finite power series) into 0. Let p be an arbitrary power series. Then, given $e > 1$, we can find a polynomial p_e such that $p - p_e \varepsilon N^{e+1}$. Then $D'\{p\} = D'\{p - p_e\} \varepsilon N^e$. Hence $D'\{p\} \varepsilon \bigcap_{e=1}^{\infty} N^e$, i. e., $D'\{p\} = 0$, whence $D = \sum_{i=1}^{n} D\{x_i\} D_i$.

If U and V are differentiations, so is $U \circ V = VU - UV$. The set of differentiations constitutes a Lie ring with the multiplication $(U, V) \rightarrow U \circ V$.

Definition 2.2. Let ω be a function defined on the s-fold set theoretical product of the module of differentiations by itself, and taking values in $K\langle x_1, \cdots, x_n \rangle$, which possesses the following properties:

(1) If $U_1, \cdots, U_s$ denote differentiations then $\omega\{U_1, \cdots, U_s\} = 0$, whenever two of the U_i's are equal.

(2) For fixed $U_2, \cdots, U_s$, the mapping $U \rightarrow \omega\{U, U_2, \cdots, U_s\}$ is an operator homomorphism of the $K\langle x_1, \cdots, x_n \rangle$-module of the differentiations into $K\langle x_1, \cdots, x_n \rangle$, regarded in the natural way as a module over itself.

Then ω is called a homogeneous differential form of degree s on $K\langle x_1, \cdots, x_n \rangle$. By a form of degree 0 we shall simply mean an element of $K\langle x_1, \cdots, x_n \rangle$.

Thus, speaking more loosely, a homogeneous differential form of degree s on $K\langle x_1, \cdots, x_n \rangle$ is an alternating s-linear function on the set of differentiations, taking values in $K\langle x_1, \cdots, x_n \rangle$.

It is easy to see that every homogeneous differential form of degree $> n$ must be 0. Clearly, a homogeneous differential form ω of degree s is determined completely by the values $\omega\{D_{i_1}, \cdots, D_{i_s}\}$, for $i_1 < \cdots < i_s$. Moreover, there always exists a differential form ω such that these values are arbitrarily prescribed elements of $K\langle x_1, \cdots, x_n \rangle$. It follows that, with the natural structure as a $K\langle x_1, \cdots, x_n \rangle$-module, the homogeneous differential forms of degree s constitute a free $K\langle x_1, \cdots, x_n \rangle$-module of rank $n!/s!(n-s)!$, for $0 \leqq s \leqq n$.

Next, we wish to define the Grassmann or 'outer' product of two differential forms. To this end, we introduce the following auxiliary functions:

Let σ be the function defined on all integers such that $\sigma\{a\} = 1$, for $a > 0$; $\sigma\{a\} = -1$, for $a < 0$; $\sigma\{a\} = 0$, for $a = 0$.

If A and B are arbitrary sets of positive integers, we set

$$e(A, B) = \prod_{i \varepsilon A, j \varepsilon B} \sigma\{j - i\},$$

(where the product of no factors is to be interpreted as 1).

Now let θ and ω be homogeneous differential forms of degree s and t, respectively. We define a function $\theta\omega$ by setting

$$(\theta\omega)\{U_1, \cdots, U_{s+t}\} = \sum_{A,B} \theta\{U_{a_1}, \cdots, U_{a_s}\}\omega\{U_{b_1}, \cdots, U_{b_t}\},$$

where the sum is to be taken over all ordered sets of integers

$$A = (a_1, \cdots, a_s),\ \text{with } 1 \leqq a_1 < \cdots < a_s \leqq s + t, \text{ and}$$
$$B = (b_1, \cdots, b_t),\ \text{with } 1 \leqq b_1 < \cdots < b_t \leqq s + t.$$

Clearly, $\theta\omega$ is $(s + t)$-linear. A straightforward computation shows that $\theta\omega$ is alternating. Thus, $\theta\omega$ is a homogeneous differential form of degree $s + t$. It can be checked directly that our outer multiplication $(\theta, \omega) \rightarrow \theta\omega$ is associative and distributive. Finally, with θ and ω as above, we have $\theta\omega = (-1)^{st}\omega\theta$.

If $p \varepsilon K\langle x_1, \cdots, x_n\rangle$ we define a homogeneous differential form dp of degree 1 by setting $(dp)\{U\} = U\{p\}$, for every differentiation U. In particular, the dx_i form a set of independent generators, over $K\langle x_1, \cdots, x_n\rangle$, for the differential forms of degree 1. Their ordered products $dx_{i_1} \cdots dx_{i_s}$ constitute a set of independent generators for the homogeneous differential forms of degree s. Explicitly, we have

$$\omega = \sum_{i_1 < \cdots < i_s} \omega\{D_{i_1}, \cdots, D_{i_s}\} dx_{i_1} \cdots dx_{i_s},$$

as can be verified easily.

We wish to extend the operator d to a mapping of homogeneous differential forms of degree s into forms of degree $s + 1$, such that

(1) $dd = 0$;

(2) $d(\theta\omega) = (d\theta)\omega + (-1)^s\theta(d\omega)$, where s is the degree of θ. This is accomplished by the coboundary formula of Eilenberg and MacLane: if ω is of degree s we define

$$(d\omega)\{U_1, \cdots, U_{s+1}\} = \sum_{i=1}^{s+1} (-1)^{i-1} U_i\{\omega\{U_1, \cdots, \hat{U}_i, \cdots, U_{s+1}\}\}$$
$$+ \sum_{p<q} (-1)^{p+q-1} \omega\{U_p \circ U_q, U_1, \cdots, \hat{U}_p, \cdots, \hat{U}_q, \cdots, U_{s+1}\},$$

where the ^ indicates omission of the argument below it. The verification of the properties (1) and (2) is rather lengthy; we shall indicate only its main outline.

First, one verifies by a direct computation that (2) holds for $s = 0$ and for $s = 1$ and that $dd\theta = 0$ if θ is of degree 0. Then one establishes (2) in general by an easy induction on the degree s of θ.

Now one verifies that $d\theta$ is a differential form, i. e. is alternating and linear, by direct computation for $s = 1$ (this is trivial for $s = 0$), and induction on s thereafter.

Finally, one establishes (1), i. e., $d(d(\theta)) = 0$, by induction on the degree of θ.

The inductions are based on the use of (2) by writing a form of degree s as a sum of products of forms of degree 1 by forms of degree $s - 1$.

An important fact for us is that the converse of (1) is true:

THEOREM 2. 2. *Let θ be a homogeneous differential form of degree $\geqq 1$ such that $d\theta = 0$. Then there exists a homogeneous form ϕ with $d\phi = \theta$.*

Proof. We may write θ in the form

$$\theta = \sum_{1 \leqq i_1 < \cdots < i_s \leqq n} \theta_{i_1 \ldots i_s} \, dx_{i_1} \cdots dx_{i_s},$$

where $\theta_{i_1 \ldots i_s} = \theta\{D_{i_1}, \cdots, D_{i_s}\}$. If ϕ is any homogeneous differential form of degree $s - 1$ we write similarly $\phi = \sum \phi_{i_1 \ldots i_{s-1}} dx_{i_1} \cdots dx_{i_{s-1}}$, and we have $d\phi = \sum d\phi_{i_1 \ldots i_{s-1}} dx_{i_1} \cdots dx_{i_{s-1}} = \sum D_j\{\phi_{i_1 \ldots i_{s-1}}\} dx_j dx_{i_1} \cdots dx_{i_{s-1}}$. The condition $d\phi = \theta$ is therefore equivalent to

$$\sum_{r=1}^{s} (-1)^{r-1} D_{i_r}\{\phi_{i_1 \ldots \hat{i}_r \ldots i_s}\} = \theta_{i_1 \ldots i_s}, \qquad 1 \leqq i_1 < \cdots < i_s \leqq n.$$

We shall prove the existence of a solution by reducing the problem to the case $s = n$. (If $s > n$ we have $\theta = 0$, and the problem is trivial). If $s = n$ we have to solve only a single equation:

$$\sum_{r=1}^{n} (-1)^{r-1} D_r\{\phi_{1 \ldots \hat{r} \ldots n}\} = \theta_{1 \ldots n}.$$

This can be solved trivially by a quadrature; for instance, we may take $\phi_{1 \ldots \hat{r} \ldots n} = 0$, for $r > 1$, and determine $\phi_{2 \ldots n}$ such that $D_{11}\{\phi_{2 \ldots n}\} = \theta_{1 \ldots n}$.

If $s < n$ then, for $1 < i_2 < \cdots < i_s \leqq n$, let us find (by quadratures) elements $\phi^*_{i_2\ldots i_s}$ in $K\langle x_1, \cdots, x_n\rangle$ such that $D_1\{\phi^*_{i_2\ldots i_s}\} = \theta_{1i_2\ldots i_s}$. We now try to determine suitable elements $\phi_{i_2\ldots i_s}$, for $1 < i_2 < \cdots < i_s \leqq n$, in the form $\phi_{i_2\ldots i_s} = \phi^*_{i_2\ldots i_s} + \psi_{i_2\ldots i_s}$, where $\psi_{i_2\ldots i_s} \varepsilon K\langle x_2, \cdots, x_n\rangle$, so that $D_1\{\psi_{i_2\ldots i_s}\} = 0$. Our conditions then reduce to the following:

$$\sum_{r=2}^{s} (-1)^{r-1} D_{i_r}\{\phi_{1i_2\ldots\hat{i}_r\ldots i_s}\} = 0, \qquad \text{for } 1 < i_2 < \cdots < i_s \leqq n,$$

and

$$\sum_{r=1}^{s} (-1)^{r-1} D_{i_r}\{\phi^*_{i_1\ldots\hat{i}_r\ldots i_s} + \psi_{i_1\ldots\hat{i}_r\ldots i_s}\} = \theta_{i_1\ldots i_s}, \text{ for } 1 < i_1 < \cdots < i_s \leqq n.$$

The first set can be met simply by taking $\phi_{1i_2\ldots i_{s-1}} = 0$, $1 < i_2 < \cdots < i_{s-1} \leqq n$. The second set may be written

$$\sum_{r=1}^{s} (-1)^{r-1} D_{i_r}\{\psi_{i_1\ldots\hat{i}_r\ldots i_s}\} = \chi_{i_1\ldots i_s},$$

where

$$\chi_{i_1\ldots i_s} = \theta_{i_1\ldots i_s} + \sum_{r=1}^{s} (-1)^r D_{i_r}\{\phi^*_{i_1\ldots\hat{i}_r\ldots i_s}\}.$$

Now we have

$$\begin{aligned} D_1\{\chi_{i_1\ldots i_s}\} &= D_1\{\theta_{i_1\ldots i_s}\} + \sum_{r=1}^{s} (-1)^r D_{i_r} D_1\{\phi^*_{i_1\ldots\hat{i}_r\ldots i_s}\} \\ &= D_1\{\theta_{i_1\ldots i_s}\} + \sum_{r=1}^{s} (-1)^r D_{i_r}\{\theta_{1i_1\ldots\hat{i}_r\ldots i_s}\} \\ &= 0, \text{ since } d\theta = 0. \end{aligned}$$

Hence $\chi_{i_1\ldots i_s} \varepsilon K\langle x_2, \cdots, x_n\rangle$. Our conditions are therefore equivalent to the relation $d\psi = \chi$, in $K\langle x_2, \cdots, x_n\rangle$.

Now consider the differential form

$$\rho = \sum_{1 < i_2 < \cdots < i_s \leqq n} \phi^*_{i_2\ldots i_s} dx_{i_2} \cdots dx_{i_s}.$$

We have

$$\begin{aligned} d\rho &= \sum D_j\{\phi_{i_2\ldots i_s}\} dx_j dx_{i_2} \cdots dx_{i_s} \\ &= \sum D_1\{\phi^*_{i_2\ldots i_s}\} dx_1 dx_{i_2} \cdots dx_{i_s} \\ &\qquad + \sum_{1 < i_1 < \cdots < i_s \leqq n} (-1)^{r-1} D_{i_r}\{\phi^*_{i_1\ldots\hat{i}_r\ldots i_s}\} dx_{i_1} \cdots dx_{i_s} \\ &= \sum \theta_{1i_2\ldots i_s} dx_1 dx_{i_2} \cdots dx_{i_s} + \sum (-1)^{r-1} D_{i_r}\{\phi^*_{i_1\ldots\hat{i}_r\ldots i_s}\} dx_{i_1} \cdots dx_{i_s}. \end{aligned}$$

Hence

$$\theta - d\rho = \sum_{1 < i_1 < \cdots < i_s \leqq n} \chi_{i_1\ldots i_s} dx_{i_1} \cdots dx_{i_s}.$$

Since this involves only the x_i and dx_i for $i > 1$, its derivative as a form χ on $K\langle x_2, \cdots, x_n\rangle$ is formally the same as its derivative as a form on $K\langle x_1, \cdots, x_n\rangle$. Since $d(\theta - d\rho) = 0$, this means that χ, regarded as a differential form on $K\langle x_2, \cdots, x_n\rangle$, has derivative 0. Thus our above conditions are of the same type as the original conditions, with the number of variables reduced to $n - 1$. By repeating this reduction we finally reach the case $s = n$, and hence obtain a solution.

It is important to observe that the coefficients $\phi_{i_1 \ldots i_{s-1}}$ of the desired form ϕ can be obtained by applying quadratures and partial differentiations to the coefficients $\theta_{i_1 \ldots i_s}$.

Let L be a Lie algebra of dimension $r \leqq n$ over K. We construct the Grassmann algebra G over L whose underlying linear space is the direct sum of the spaces G_s of the s-linear alternating functions on L, with values in K, for $s = 0, 1, \cdots, r$, and where $G_0 = K$. The multiplication in G is defined exactly like the outer multiplication of differential forms. We also define a K-linear mapping δ which maps each G_s into G_{s+1}, as follows: if $g \in G_s$ $(s > 0)$ and u_i are elements of L we set

$$(\delta g)\{u_1, \cdots, u_{s+1}\} = \sum_{p<q} (-1)^{p+q-1} g\{u_p \circ u_q, u_1, \cdots, \hat{u}_p, \cdots, \hat{u}_q, \cdots, u_{s+1}\}.$$

We define $\delta(G_0) = (0)$, for completeness.

Exactly as in the case of the operator d for differential forms, we can show that $\delta(gh) = (\delta g)h + (-1)^s g(\delta h)$ and hence that $\delta g \in G_{s+1}$, and $\delta\delta = 0$.

It is no longer true here that $\delta g = 0$ implies that $g = \delta h$, with some $h \in G_{s-1}$. The additive group of the elements of G_s which are mapped into 0 by δ, modulo $\delta(G_{s-1})$ is an important invariant of L, its s-dimensional cohomology group $H^s(L)$. (See [6]).

We shall later construct an isomorphism of G into the Grassmann ring Ω of the differential forms on $K\langle x_1, \cdots, x_n\rangle$. In order that this should lead to a representation of L it is necessary that certain regularity conditions be satisfied. We shall proceed to consider this question in detail.

Denote by Ω_i the module of the homogeneous differential forms of degree i on $K\langle x_1, \cdots, x_n\rangle$. A mapping α of G into Ω will be called an isomorphism if the following conditions are satisfied:

(1) On $G_0 = K$, α is the natural injection of K into $K\langle x_1, \cdots, x_n\rangle$.

(2) For each i, α maps G_i isomorphically into Ω_i.

(3) $\alpha(gg') = \alpha(g)\alpha(g')$, and $\alpha(\delta g) = d\alpha(g)$, for all $g, g' \in G$.

Since, for $s > 0$, $G_s = (G_1)^s$, it follows from (3) that an isomorphism α is completely determined by its restriction to G_1. The regularity condition which we wish to impose is the following: denote by Q_r the submodule of Ω_r consisting of all forms ω for which $\omega\{U_1, \cdots, U_r\} \in N$, the ideal of non-units, for all differentiations U_i. Then we shall say that the isomorphism α is regular if $\alpha(G_1)^r \not\subseteq Q_r$.

Let $g_1, \cdots, g_r$ be any basis for G_1 over K. Then $\alpha(G_1)^r$ consists of all the K-multiples of the single element $\alpha(g_1) \cdots \alpha(g_r)$, and our regularity condition is equivalent to the condition $\alpha(g_1) \cdots \alpha(g_r) \notin Q_r$. Each $\alpha(g_i)$ may be written in the form $\alpha(g_i) = \sum_{j=1}^{n} a_{ij} dx_j$, where the $a_{ij} \in K\langle x_1, \cdots, x_n\rangle$, and we have $\alpha(g_1) \cdots \alpha(g_r) = \sum | a_{ij} |_{j \in A} dx_{\alpha_1} \cdots dx_{\alpha_r}$, where the sum is taken over all sets A: $1 \leqq \alpha_1 < \cdots < \alpha_r \leqq n$. Hence, if α is regular, there exists at least one such set A for which the determinant $| a_{ij} |_{j \in A} \notin N$. If $B = (\beta_1, \cdots, \beta_{n-r})$ is the complementary set of indices this evidently gives $dx_{\alpha_i} = \sum_{j=1}^{r} b_{ij}\alpha(g_j) + \sum_{k=1}^{n-r} c_{ik} dx_{\beta_k}$, with $b_{ij}, c_{ik} \in K\langle x_1, \cdots, x_n\rangle$. Hence the elements $\alpha(g_1), \cdots, \alpha(g_r)$; dx_{β_k} constitute a system of generators for Ω_1. This system is independent since

$$\alpha(g_1) \cdots \alpha(g_r) dx_{\beta_1} \cdots dx_{\beta_{n-r}} = e(A, B) | a_{ij} |_{j \in A} dx_1 \cdots dx_n \neq 0.$$

We are now in a position to prove the following theorem:

THEOREM 2.3. *Let L be a Lie algebra of dimension r over K, and let G be the associated Grassmann algebra. Suppose there is given a regular isomorphism $g \to \bar{g}$ of G into the module Ω of the differential forms on $K\langle x_1, \cdots, x_n\rangle$. Then there is an isomorphism $u \to \bar{u}$ of L onto a Lie algebra of differentiations in $K\langle x_1, \cdots, x_n\rangle$, such that $g\{u\} = \bar{g}\{\bar{u}\}$ for every $g \in G$ and $u \in L$.*

Proof. Let $g_1, \cdots, g_r$ be a basis for G_1 over K. By the above, we may suppose that $\bar{g}_1, \cdots, \bar{g}_r, dx_{r+1}, \cdots, dx_n$ form a free system of generators for Ω_1 over $K\langle x_1, \cdots, x_n\rangle$. Then it is clear that, for each $u \in L$, there exists a unique differentiation $\bar{u}$ in $K\langle x_1, \cdots, x_n\rangle$, such that $(dx_j)\{\bar{u}\} = 0$ for $j > r$, and $\bar{g}_i\{\bar{u}\} = g_i\{u\}$.

We shall prove that the mapping $u \to \bar{u}$ is an isomorphism. Since it is evidently a K-linear isomorphism of the vector space L into the module of differentiations, there remains only to show that $\overline{u \circ v} = \bar{u} \circ \bar{v}$. Now for $g \in G_1$ we have $(d\bar{g})\{\bar{u}, \bar{v}\} = \bar{u}\{\bar{g}\{\bar{v}\}\} - \bar{v}\{\bar{g}\{\bar{u}\}\} + \bar{g}\{\bar{u} \circ \bar{v}\} = \bar{g}\{\bar{u} \circ \bar{v}\}$, since $\bar{g}\{\bar{u}\} = g\{u\} \in K$, and $\bar{g}\{\bar{v}\} = g\{v\} \in K$. Write $\delta g = \sum_{p<q} a_{pq} g_p g_q$, where $a_{pq} \in K$.

Then

$$
\begin{aligned}
(\delta g)\{u,v\} &= \sum_{p<q} a_{pq}(g_p\{u\}g_q\{v\} - g_p\{v\}g_q\{u\}) \\
&= \sum_{p<q} a_{pq}(\bar{g}_p\{\bar{u}\}\bar{g}_q\{\bar{v}\} - \bar{g}_p\{\bar{v}\}\bar{g}_q\{\bar{u}\}) = \overline{\delta g}\{\bar{u},\bar{v}\} = (d\bar{g})\{\bar{u},\bar{v}\}
\end{aligned}
$$

i. e., $g\{u \circ v\} = \bar{g}\{\bar{u} \circ \bar{v}\}$, i. e., $\bar{g}\{\overline{u \circ v}\} = \bar{g}\{\bar{u} \circ \bar{v}\}$, which shows that $\overline{u \circ v} = \bar{u} \circ \bar{v}$.

3. Representation of solvable Lie algebras.

We shall base our computations on the following known theorem:

THEOREM 3. 1. *Let L be a solvable Lie algebra over a field F of characteristic 0. Then there exists a finite algebraic extension field K of F such that the extension L^0 of L over K has a composition series of the following sort: $L^0 = L_0{}^0 \supset L_1{}^0 \supset \cdots \supset L_n{}^0 = (0)$, where $L_i{}^0$ is an ideal in L^0 and is of dimension $n - i$ over K. Moreover, there is an index r such that $L_r{}^0$ coincides with the maximal nilpotent ideal of L^0.*

Proof. Let F_1 be the algebraic closure of F. Denote by L^1 the extension of L over F_1, and by N^1 the extension over F_1 of the maximal nilpotent ideal N of L. Since N^1 is an ideal in L^1, we may regard it as an L^1-module in the natural fashion. Let $N^1 = Q_0 \supset Q_1 \supset \cdots \supset Q_s = (0)$ be any composition series of this module. Then the quotients Q_{i-1}/Q_i are simple representation spaces for L^1. By theorem 1. 3, every element of L^1 induces a transformation in Q_{i-1}/Q_i which commutes with every other transformation belonging to our representation. Making use of Schur's lemma and the fact that F_1 is algebraically closed, we conclude that all these transformations are scalar multiplications. Since Q_{i-1}/Q_i is simple, this implies that it is of dimension 1 over F_1. This means that the ideals Q_i are of dimension $s - i$ over F_1. We choose elements $\nu_1, \cdots, \nu_s$ such that $\nu_i \varepsilon Q_{i-1}$ but $\nu_i \not\varepsilon Q_i$. Then this set constitutes a basis for N^1 over F_1. Each ν_i may be expressed in terms of a basis for L over F, allowing the coefficients to lie in F_1. This finite set of coefficients generates a finite algebraic extension K of F.

Let N^0 be the extension of N over K. If $L^0{}_{n-i}$ is the vector space over K which is spanned by $\nu_s, \nu_{s-1}, \cdots, \nu_{s-i+1}$ then these $L^0{}_j$ are ideals in L^0 and $L^0{}_{n-s} = N^0$. The maximal nilpotent ideal of L^0 must evidently contain N^0. Since $L^0 \circ L^0 = (L \circ L)^0 \subset N^0$, every subspace of L^0 which contains N^0 is an ideal, and we can trivially determine $L_1{}^0, \cdots, L^0{}_{n-s-1}$ so as to satisfy the conditions of Theorem 3.

Henceforth we shall suppose that L is a solvable Lie algebra over K which already possesses a composition series

$$L = L_0 \supset L_1 \supset \cdots \supset L_r = N \supset L_{r+1} \supset \cdots \supset L_n = (0),$$

where each L_i is an ideal in L and is of dimension $n - i$ over K. We select a basis $u_1, \cdots, u_n$ of L over K such that $u_i \in L_{i-1}$, $u_i \notin L_i$. If we write $u_i \circ u_j = \sum_{k=1}^{n} c_{ijk} u_k$, the $c_{ijk} \in K$ have the following properties:

(1) $c_{ijk} = 0$, unless $k \geqq i$, $k \geqq j$, and $k > r$.

(2) If $i > r$ and $j > r$ then $c_{ijk} = 0$, unless $k > i$ and $k > j$.

In fact, (1) follows from the construction of the u_i and the fact that $L \circ L \subseteq N$, the maximal nilpotent ideal. (2) follows from the fact that $u_i \in N$ for $i > r$ and that N is nilpotent.

Let $g_1, \cdots, g_n$ be the basis for the elements of degree 1 in the Grassmann algebra over L for which $g_i\{u_j\} = \delta_{ij}$. Then we have $(\delta g_k)\{u_i, u_j\} = g_k\{u_i \circ u_j\} = c_{ijk}$, whence, using (1), $\delta g_k = \sum_{i < j \leqq k} c_{ijk} g_i g_j$.

We wish to construct a regular isomorphism $g \rightarrow \bar{g}$ of the Grassmann algebra G over L into Ω, the module of the homogeneous differential forms on $K\langle x_1, \cdots, x_n\rangle$. We shall do this by constructing successively $\bar{g}_1 = \omega_1, \cdots, \bar{g}_n = \omega_n$ so as to satisfy all the conditions laid down in **2**.

First, we note that, for $k \leqq r$, we have $\delta g_k = 0$. Therefore, we may define $\omega_k = dx_k$, for $k \leqq r$.

We shall prove that we can find forms ω_k such that $d\omega_k = \sum_{i < j \leqq k} c_{ijk} \omega_i \omega_j$ and that, moreover, they may be taken in the form

$$\omega_k = f_k dx_k + \sum_{s=1}^{k-1} P_{ks}(f_1, \cdots, f_{k-1}; x_1, \cdots, x_{k-1}) dx_s,$$

where the P_{ks} are polynomials, and where $f_i = \exp\left(\sum_{p=1}^{r} c_{pii} x_p\right)$. Note that, by (1), $f_i = 1$ for $i \leqq r$, so that our assertion holds true for $\omega_1, \cdots, \omega_r$.

Now let $k > r$, and suppose that $\omega_1, \cdots, \omega_{k-1}$ have already been determined. The above differential equation for ω_k may be written:

$$d\omega_k - \left(\sum_{i=1}^{k-1} c_{ikk} \omega_i\right) \omega_k = \sum_{i < j < k} c_{ijk} \omega_i \omega_j.$$

By the property (2) of the c_{ijk} we see that the sum on the left has non-zero terms only up to $i = r$. Hence the ω_i on the left are the dx_i. If we multiply by f^{-1}_k the equation takes the form

$$d(f^{-1}_k\omega_k) = f^{-1}_k \sum_{i<j<k} c_{ijk}\omega_i\omega_j = \sigma_k,$$

say.

We wish to verify that $d\sigma_k = 0$. Direct computation gives

$$d\sigma_k = f^{-1}_k\Big[d\Big(\sum_{i<j<k} c_{ijk}\omega_i\omega_j\Big) - \sum_{p=1}^{r} c_{pkk}\omega_p \sum_{i<j<k} c_{ijk}\omega_i\omega_j\Big].$$

On the other hand, since $\delta(\delta g_k) = 0$, we have

$$\begin{aligned}\delta\Big(\sum_{i<j<k} c_{ijk}g_ig_j\Big) &= -\delta\Big(\sum_{p=1}^{r} c_{pkk}g_pg_k\Big)\\ &= -\sum_{p=1}^{r} c_{pkk}(\delta g_p)g_k + \sum_{p=1}^{r} c_{pkk}g_p\delta g_k\\ &= \sum_{p=1}^{r} c_{pkk}(\delta g_p)g_k + \sum_{p=1}^{r} c_{pkk}g_p\Big(\sum_{i=1}^{r} c_{ikk}g_ig_k + \sum_{i<j<k} c_{ijk}g_ig_j\Big).\end{aligned}$$

Noting that $\delta g_p = 0$ for $p \leqq r$, and that the square of a homogeneous element of degree 1 in a Grassmann algebra is 0, we see that this relation reduces to

$$\delta\Big(\sum_{i<j<k} c_{ijk}g_ig_j\Big) = \sum_{p=1}^{r} c_{pkk}g_p \sum_{i<j<k} c_{ijk}g_ig_j.$$

Since, by our inductive hypothesis, the ω_i must satisfy the same relations over K as the g_i for $i = 1, \cdots, k-1$, the last relation implies that $d\sigma_k = 0$.

Now we may apply Theorem 2. 2 and conclude that there exists a homogeneous differential form ϕ_k of degree 1 such that $d\phi_k = \sigma_k$. Since the coefficients of ϕ_k can be obtained by applying partial differentiations and quadratures to the coefficients of σ_k, it follows from our inductive assumption about the form of the ω_i for $i < k$ that ϕ_k may be taken of the form $\phi_k = f^{-1}_k \sum_{s=1}^{k-1} P_{ks}(f_1, \cdots, f_{k-1}; x_1, \cdots, x_{k-1})\,dx_s$, where the P_{ks} are polynomials. We set $\omega_k = f_k(dx_k + \phi_k)$, i. e.,

$$\omega_k = f_k dx_k + \sum_{s=1}^{k-1} P_{ks}(f_1, \cdots, f_{k-1}; x_1, \cdots, x_{k-1})\,dx_s.$$

Then it can be verified quite easily that $d\omega_k = \sum_{i<j\leqq k} c_{ijk}\omega_i\omega_j$.

Hence we can determine $\omega_1, \cdots, \omega_n$ so as to satisfy the above conditions. Furthermore, we have then

$$\omega_1 \cdots \omega_n = \prod_{k=1}^{n} \Big(f_k dx_k + \sum_{s=1}^{k-1} P_{ks}dx_s\Big) = \Big(\prod_{k=1}^{n} f_k\Big)\,dx_1 \cdots dx_n.$$

Since $\prod_{k=1}^{n} f_k = \exp\left(\sum c_{pkk} x_p\right)$ is a unit, we see that the correspondence $g_i \to \omega_i$ can be extended to a regular isomorphism of G into Ω. By Theorem 2. 3, there is an isomorphism $u \to \bar{u}$ of L onto a Lie algebra of differentiations in $K\langle x_1, \cdots, x_n\rangle$ such that $g_i\{u\} = \omega_i\{\bar{u}\}$. From the form of the ω_i we conclude that

(1) $\bar{u}\{x_i\} \in K$, for $i \leqq r$;

(2) $\bar{u}\{x_j\}$ is a polynomial in $f_1, \cdots, f_j; f^{-1}{}_1, \cdots, f^{-1}{}_j; x_1, \cdots, x_{j-1}$;

(3) If $j > r$ and $i \leqq r$ then $\bar{u}_j\{x_i\} = 0$.

It follows that

(4) $\bar{u}\{f_k\} = af_k$, with $a \in K$;

(5) If $j > r$ then $\bar{u}_j\{f_k\} = 0$.

We wish to show next that repeated application of the $\bar{u}$ to the x_j generates only a finite dimensional K-linear subspace of $K\langle x_1, \cdots, x_n\rangle$. It is clear that every element of $K\langle x_1, \cdots, x_n\rangle$ which can be obtained from the x_i by repeatedly applying operators $\bar{u}$ may be written in the form $\sum_{a} f_1^{e_1(a)} \cdots f_n^{e_n(a)} P_a(x_1, \cdots, x_n)$, where the P_a are polynomials and the $e_i(\alpha)$ are positive or negative integers. There are at most n^2 linearly independent first transforms $\bar{u}\{x_i\}$. Let $M > 1$ be an upper bound for the total degrees of the P_a which occur in these transforms. We introduce a weight function w for polynomials $P(x_1, \cdots, x_n)$ as follows:

If m_i is the degree of $P(x_1, \cdots, x_n)$ in x_i we define $w(P) = \sum_{i=1}^{n} m_i M^{2i}$. Then, if P is of degree $m_i \neq 0$ in x_i, we have $D_i\{P\}$ of degree $m_i - 1$ in x_i. This shows that $\bar{u}\{P\} = \sum_{i=1}^{n} D_i\{P\}\bar{u}\{x_i\}$ can be written in the above form with all the polynomials P_a of weight no greater than

$$\mathrm{Max}_i[w(P) - M^{2i} + M \cdot M^{2(i-1)}] < w(P).$$

Now we have

$$\bar{u}\{\sum_{a} f_1^{e_1(a)} \cdots f_n^{e_n(a)} P_a\} = \sum_{a} f_1^{e_1(a)} \cdots f_n^{e_n(a)} a(\alpha) P_a + \sum_{a} f_1^{e_1(a)} \cdots f_n^{e_n(a)} \bar{u}\{P_a\},$$

where the $a(\alpha) \in K$. If the second sum is written in the standard form, the exponents $e_i(\alpha)$ may be increased numerically, but the weights of all the new polynomials which appear are less than the maximum of the $w(P_a)$.

It follows that all the repeated transforms of the x_i may be written in the standard form in such a way that there is an upper bound for all the $|e_i(\alpha)|$ which appear, and such that each polynomial P_α is of weight no greater than $w(x_n) = M^{2n}$. Evidently, all these elements are contained in a finite dimensional K-linear subspace of $K\langle x_1, \cdots, x_n\rangle$. Furthermore, it follows from the above and from (5) that $\bar{N}$ consists entirely of nilpotent differentiations. We may state our result as follows:

THEOREM 3. 2. *Let L be a solvable Lie algebra over a field F of characteristic 0. Then there exists a finite algebraic extension K of F and an isomorphism $u \rightarrow \bar{u}$ of L onto a Lie algebra $\bar{L}$ of differentiations in $K\langle x_1, \cdots, x_n\rangle$, where n is the dimension of L over F. There is a finite dimensional F-linear subspace U of $K\langle x_1, \cdots, x_n\rangle$ which contains the x_i and is such that $\bar{L}\{U\} \subseteq U$. If $\bar{\bar{u}}$ is the restriction to U of $\bar{u}$, the mapping $u \rightarrow \bar{\bar{u}}$ is an isomorphism of L onto a Lie algebra $\bar{\bar{L}}$ of linear transformations in U. Moreover, if N is the maximal nilpotent ideal of L then $\bar{N}$ consists entirely of nilpotent differentiations, and hence $\bar{\bar{N}}$ consists entirely of nilpotent transformations.*

4. Adjunction of derivations. Let L be a solvable Lie algebra and suppose that the elements of L are identified with the differentiations in $K\langle x_1, \cdots, x_n\rangle$ which were obtained in **3**. Let τ be any derivation in L and write $\tau\{u_i\} = \sum_{j=1}^{n} a_{ij}u_j$. We wish to find a differentiation t in $K\langle x_1, \cdots, x_n\rangle$ such that $tu - ut = \tau\{u\}$ for every $u \varepsilon L$.

We may assume that t is of the form $t = \sum_{i=1}^{n} g_i u_i$, with $g_i \varepsilon K\langle x_1, \cdots, x_n\rangle$, and we have to determine the g_i so as to satisfy the relations $tu_j - u_jt = \sum_{k=1}^{n} a_{jk}u_k$, i. e.,

$$\sum_{i=1}^{n} g_i(u_iu_j - u_ju_i) - \sum_{i=1}^{n} u_j\{g_i\}u_i = \sum_{k=1}^{n} a_{jk}u_k,$$

i. e., $-\sum_{i=1}^{n} g_ic_{ijk} - u_j\{g_k\} = a_{jk}$ or, using the property (1), **3**, of the c_{ijk},

$$u_j\{g_k\} = -\left(a_{jk} + \sum_{i=1}^{k} g_ic_{ijk}\right). \tag{1}$$

Let $\omega_1, \cdots, \omega_n$ be the differential forms of degree 1 on $K\langle x_1, \cdots, x_n\rangle$ for which $\omega_i\{u_j\} = \delta_{ij}$. Then the last relations give

$$dg_k = -\sum_{j=1}^{n} (a_{jk} + \sum_{i=1}^{k} g_i c_{ijk}) \omega_j.$$

This may be written

$$dg_k - \sum_{j=1}^{r} g_k c_{jkk} \omega_j = -\sum_{j=1}^{n} (a_{jk} + \sum_{i<k} g_i c_{ijk}) \omega_j.$$

Since $\omega_j = dx_j$ for $j \leqq r$, this gives

$$(2) \qquad dh_k = -f^{-1}{}_k \sum_{j=1}^{n} (a_{jk} + \sum_{i<k} g_i c_{ijk}) \omega_j$$

where $h_k = f^{-1}{}_k g_k$. For $k \leqq r$ our equations reduce to $dh_k = -\sum_{j=1}^{n} a_{jk} \omega_j$, i. e., $dg_k = -\sum_{j=1}^{n} a_{jk} \omega_j$. By Theorem 1. 5, $\tau\{N\} \subseteqq N$, whence $a_{jk} = 0$ if $j > r$ and $k \leqq r$. Therefore, the last equations reduce to $dg_k = -\sum_{j=1}^{r} a_{jk} dx_j$, which can be satisfied with $g_k = -\sum_{j=1}^{r} a_{jk} x_j$, for $k \leqq r$.

Suppose that we have already found $g_1, \cdots, g_{s-1}$ (and hence $h_1, \cdots, h_{s-1}$) such that (1) (and therefore (2)) hold for all $k \leqq s-1$. Set

$$\rho_s = f^{-1}{}_s \sum_{j=1}^{n} (a_{js} + \sum_{i<s} g_i c_{ijs}) \omega_j.$$

We wish to show that $d\rho_s = 0$. We shall do this by proving that, for all p, q, $u_q\{\rho_s\{u_p\}\} - u_p\{\rho_s\{u_q\}\} = \rho_s\{u_p \circ u_q\}$.

Now $\rho_s\{u_p\} = f^{-1}{}_s(a_{ps} + \sum_{i<s} g_i c_{ips})$. A straightforward computation using (1) and the properties of the c_{ijk} gives

$$f_s(u_q\{\rho_s\{u_p\}\} - u_p\{\rho_s\{u_q\}\}) = \sum_{i=1}^{n} (a_{pi} c_{iqs} - a_{qi} c_{ips}) + \sum_{j=1}^{s-1} g_j \sum_{i=1}^{n} (c_{jpi} c_{iqs} - c_{jqi} c_{ips}).$$

Since

$$\tau\{u_p\} \circ u_q - \tau\{u_q\} \circ u_p = \tau\{u_p \circ u_q\} = \sum_{i=1}^{n} c_{pqi} \tau\{u_i\},$$

we have

$$\sum_{i=1}^{n} (a_{pi} c_{iqs} - a_{qi} c_{ips}) = \sum_{i=1}^{n} c_{pqi} a_{is}.$$

Since

$$(u_j \circ u_p) \circ u_q - (u_j \circ u_q) \circ u_p = u_j \circ (u_p \circ u_q),$$

we have

$$\sum_{i=1}^{n} (c_{jpi}c_{iqs} - c_{jqi}c_{ips}) = \sum_{i=1}^{n} c_{pqi}c_{jis}.$$

Hence

$$f_s(u_q\{\rho_s\{u_p\}\} - u_p\{\rho_s\{u_q\}\}) = \sum_{i=1}^{n} c_{pqi}(a_{is} + \sum_{j<s} g_j c_{jis}) = f_s\rho_s\{u_p \circ u_q\},$$

which proves that $d\rho_s = 0$.

By Theorem 2. 2, there exists an element $h_s \varepsilon K\langle x_1, \cdots, x_n\rangle$ such that $dh_s = -\rho_s$. If we put $g_s = f_s h_s$ we see that g_s satisfies (1). Thus, the required differentiation exists.

Next we wish to investigate the effect of applying t to the elements of the space U of Theorem 3. 2. An important fact is that $t\{f_k\} = 0$, for all k. We have $t\{f_k\} = f_k \sum_{p=1} t\{x_p\} c_{pkk}$. Since, for $p \leqq r$, $t\{x_p\} = g_p = -\sum_{j=1}^{r} a_{jp}x_j$, we obtain

$t\{f_k\} = -f_k \sum_{j=1}^{r} (\sum_{p=1}^{r} a_{jp}c_{pkk}) x_j$. We shall show that $\sum_{p=1}^{r} a_{jp}c_{pkk} = 0$.

If $u = \sum_{i=1}^{n} e_i u_i$ we have

$$u_k \circ u = \sum_{i=1}^{n} e_i \sum_{q=1}^{n} c_{kiq}u_q = -\sum_{i=1}^{n} e_i c_{ikk} u_k - \sum_{q>k} (\sum_{i=1}^{n} e_i c_{ikq}) u_q,$$

which shows that $-\sum_{i=1}^{n} e_i c_{ikk}$ is a characteristic root of D_u for every k. Hence if $u \varepsilon N$ then $\sum_{i=1}^{n} e_i c_{ikk} = 0$. By Theorem 1. 5., $\tau(u_j) \varepsilon N$. We may therefore conclude that $\sum_{r=1}^{r} a_{jp}c_{pkk} = 0$, whence $t\{f_k\} = 0$.

We have seen in **3** that the elements of U can be written

$$\phi = \sum_{a} f_1^{e_1(a)} \cdots f_n^{e_n(a)} P_a(x_1, \cdots, x_n).$$

Hence

$$t\{\phi\} = \sum_{a} f_1^{e_1(a)} \cdots f_n^{e_n(a)} \sum_{i=1}^{n} g_i u_i\{P_a\}.$$

The elements $u_i\{P_a\}$ can be written as sums of products of monomials in the f_j (allowing negative exponents) by polynomials in the x_k of lower weight than P_a.

For greater convenience, we introduce the following notion: if ϕ is any element of the above form, not necessarily contained in U, we shall define the weight of ϕ as the minimum, for all standard representations, of the greatest occurring $w(P_\alpha)$. Any polynomial in the f_i, f^{-1}_i and x_i will henceforth be called a standard element.

Now we may express the result we have just obtained by saying that if ϕ is any standard element then $t\{\phi\} = \sum_{i=1}^{n} g_i\phi_i$, where the ϕ_i are standard elements of lower weight than ϕ.

Let us observe further that, by (1), we have $t\{g_k\} = -\sum_{i=1}^{n} g_i a_{ik}$. If $\phi = \sum_\gamma G_\gamma H_\gamma$, where the G_γ are polynomials in the g_i and the H_γ are standard elements, we have $t\{\phi\} = \sum_\gamma t\{G_\gamma\} H_\gamma + \sum_{\gamma,i}(G_\gamma g_i)H'_{\gamma i}$. The degree of $t\{G_\gamma\}$ is no greater than the degree of G_γ and the weight of each $H'_{\gamma i}$ is less than the weight of H_γ. Since t maps elements of weight 0 into 0, we conclude that all the repeated transforms of a standard element by t lie in a finite-dimensional K-linear subspace of $K\langle x_1, \cdots, x_n\rangle$. It follows that there exists a finite dimensional K-linear subspace $U_1 \supseteq U$ of $K\langle x\ , \cdots, x_n\rangle$ which is mapped into itself by t, and which is spanned by repeated t-transforms of elements of U.

Since $ut = tu - \tau\{u\}$, it is easily seen that the K-linear subspace spanned by repeated transforms of the elements of U by all the operations from (L, t) coincides with the K-linear subspace spanned by the repeated t-transforms of the elements of U, i. e., with U_1.

Finally, we note that it follows inductively from (1) that all the g_k are standard elements, and hence that U_1 consists entirely of standard elements. We may state our result as follows:

THEOREM 4.1. *Let L be a Lie algebra over K whose elements are differentiations in $K\langle x_1, \cdots, x_n\rangle$ with the properties listed in* **3**. *Let τ be an arbitrary derivation of L. Then there exists a differentiation t in $K\langle x_1, \cdots, x_n\rangle$ such that $tu - ut = \tau\{u\}$, for all $u \in L$. Furthermore, if V is any finite dimensional K-linear subspace of $K\langle x_1, \cdots, x_n\rangle$ which consists entirely of standard elements, there exists a finite dimensional subspace $V_1 \supseteq V$ of $K\langle x_1, \cdots, x_n\rangle$ such that all the operations from (L, t) map V_1 into itself, and V_1 consists entirely of standard elements.*

Only the last statement requires comment: we merely note that the proof of Theorem 3.2 shows that the operations from L generate from V

only a finite dimensional subspace consisting of standard elements. The argument given above now suffices to establish Theorem 4. 1.

A differentiation t in $K\langle x_1, \cdots, x_n\rangle$ will be called singular if (1) $tu - ut = 0$, for all $u \in L$, and (2) $t\{x_i\} = 0$ for $i \leqq r$.

We wish to show that the set of singular differentiations is of finite dimension over K. For a singular differentiation $t = \sum_{i=1}^{n} g_i u_i$ the differential equations (2) become

$$dh_k = -f^{-1}{}_k \sum_{j=1}^{n} \sum_{i<k} g_i c_{ijk} \omega_j.$$

If $g_1, \cdots, g_s = 0$ these equations are satisfied for $k \leqq s$. We must have $dh_{s+1} = 0$, whence $g_{s+1} = a_{s+1} f_{s+1}$; $a_{s+1} \in K$. It follows from our proof of Theorem 4. 1 that for each $i > r$ we can find a singular differentiation t_i of the form $t_i = f_i u_i + \sum_{j>i} g_{ij} u_j$.

If $s \geqq r$ the differentiation $t - a_{s+1} t_{s+1}$ is singular, and if $t - a_{s+1} t_{s+1} = \sum g'_i u_i$ we have $g'_i = 0$, for all $i \leqq s + 1$. It follows that every singular differentiation is a K-linear combination of $t_{r+1}, \cdots, t_n$.

5. Representation of arbitrary Lie algebras. Let H be an arbitrary Lie algebra over the field F of characteristic 0. Denote the maximal solvable ideal of H by L. By Theorem 1. 2, we have a linearly direct decomposition $H = T + P + L$, where T is (0) or a semisimple ideal of H such that $T \circ (P + L) = (0)$, and where P is (0) or a semisimple subalgebra of H such that if $0 \neq p \in P$ the derivation $u \rightarrow u \circ p$ is not an inner derivation in L. Since a semisimple Lie algebra has, trivially, a faithful linear representation (e. g. its adjoint representation), we may confine ourselves to the case where $T = (0)$, i. e. $H = P + L$, and there remains only the case where $P \neq (0)$.

Let $p_1, \cdots, p_m$ be a basis for P over F. By Theorem 3. 2, we may suppose that L is identified with a Lie algebra of differentiations in $K\langle x_1, \cdots, x_n\rangle$ as in **3**, **4**, where K is a finite algebraic extension of F. By Theorem 4. 1, we can find differentiations t_i in $K\langle x_1, \cdots, x_n\rangle$ such that $t_i u - u t_i = u \circ p_i$ for every $u \in L$. For $p = \sum_{i=1}^{m} a_i p_i$, $a_i \in K$, we set $p^* = \sum_{i=1}^{m} a_i t_i$, so that $p^*_i = t_i$.

Now we claim that the p^* for $p \in P$ generate a finite dimensional Lie algebra over K (and hence also a finite dimensional Lie algebra over F) with the commutation $(p^*, q^*) \rightarrow p^* \circ q^* = q^* p^* - p^* q^*$. In order to prove this, we note first that the differentiations $p^* \circ q^* - (p \circ q)^*$, for $p, q \in P$, are singular. In fact, from the relations $p^*u - up^* = u \circ p$ it follows at once that each $p^* \circ q^* - (p \circ q)^*$ commutes with every $u \in L$. Furthermore, for $k \leqq r$, we have $p^*\{x_k\} = -\sum_{s=1}^{r} a_{sk}x_s$, where the coefficients $a_{sk} \in K$ are determined from the defining relations $u_s \circ p = \sum_{k=1}^{n} a_{sk}u_k$ in L. The derivation $u \rightarrow u \circ p$ induces in L/N (N the maximal nilpotent ideal) a derivation $\bar{p}$ such that—if $\bar{u}$ denotes the coset mod N of $u \in L$—we have $\bar{p}\{\bar{u}\} = \sum_{k=1}^{r} a_{sk}\bar{u}_k$. Let p, q be elements of P and denote the matrices of $\bar{p}$, $\bar{q}$ with respect to the basis $\bar{u}_1, \cdots, \bar{u}_r$ of L/N by A and B, respectively. Then the matrices p^* and q^* with respect to the coordinates $x_1, \cdots, x_r$ are the negative transposes $-A'$ and $-B'$, respectively. The transformation $\overline{p \circ q}$ has then the matrix $AB - BA$ since it is the same as $\bar{q}\bar{p} - \bar{p}\bar{q}$. Hence the transformation of the x's by $(p \circ q)^*$ has the matrix $A'B' - B'A'$, whence we see that this is the same transformation as the transformation by $p^* \circ q^* = q^*p^* - p^*q^*$. Hence $(p^* \circ q^*)\{x_k\} = (p \circ q)^*\{x_k\}$, whence $p^* \circ q^* - (p \circ q)^*$ is singular.

Furthermore, if t is singular and q^* is an arbitrary element of the Lie algebra generated by the p^* with $p \in P$ then $t \circ q^*$ is also singular. In fact, it is evident that $(t \circ q^*)\{x_i\} = 0$, for $i \leqq r$. Also, if $u \in L$ we have $q^* \circ u \in L$, and therefore $u \circ (t \circ q^*) = -t \circ (q^* \circ u) - q^* \circ (u \circ t) = 0$.

Let R denote the Lie algebra which is generated over F by the elements p^* with $p \in P$. By the above, every element of R is a sum of a p^* and a singular differentiation. Since the set of singular differentiations is of finite dimensions over F, we conclude that R is of finite dimension over F. Let C denote the set of all singular differentiations which belong to R. We have seen that C is an ideal in R, and that $R = P + C$. Moreover, this sum is linearly direct. For, if $p^* + c = p^*_0 + c_0$ then $(p - p_0)^* = c_0 - c$ commutes with every u in L, which implies that, in H, we have $u \circ (p - p_0) = 0$, for every $u \in L$. But this implies that $p = p_0$, because of the property of P in H. Hence every element $v \in R$ can be written in the form $v = p^* + c$, where p, not only p^*, is unique. We define the mapping π of R onto P by setting $\pi\{v\} = p$. It is clear that π is a homomorphism of the Lie algebra R onto the Lie algebra P. By Theorem 1.1, there exists an isomorphism α of P into R such that $\pi\alpha$ is the identity mapping on P. Since the kernel of π is C, we have $p^* - \alpha\{p\} \in C$, whence $u \circ \alpha\{p\} = u \circ p^* = u \circ p$, for every $u \in L$.

If $h = p + u \in H$ we define $\beta\{h\} = \alpha\{p\} + u$. Then the mapping β is evidently a homomorphism. Moreover, if $\beta\{h\} = 0$ we have $\alpha\{p\} = -u$, hence $u_0 \circ p = u_0 \circ (-u)$, for every $u_0 \varepsilon L$, which, by the property of P, implies that $p = 0$. Thus, β is an isomorphism.

The differentiations $\beta\{p_i\}$ differ from the t_i only by singular differentiations, and their coefficients are still standard elements. Therefore, all the considerations made in **4** apply also to the $\beta\{p_i\}$. In particular, Theorem 4.1 holds for each $\beta\{p_i\}$. We shall write q_i for $\beta\{p_i\}$.

Now there is a finite dimensional K-linear subspace U of $K\langle x_1, \cdots, x_n\rangle$ which contains the x_i, consists of standard elements, and is mapped into itself by every $u \varepsilon L$. By repeatedly applying q_1 to the elements of U we obtain a finite dimensional K-linear subspace $U_1 \supseteq U$ which consists of standard elements and which is mapped into itself by every $u \varepsilon L$ and by q_1. Indeed, this is the second assertion of Theorem 4.1. If U_i has already been constructed, we obtain $U_{i+1} \supseteq U_i$ by repeated applications of q_{i+1}. The last of these spaces, U_m, is mapped into itself by every $u \varepsilon L$ and by q_m. We claim that, actually, U_m is also mapped into itself by $q_1, \cdots, q_{m-1}$, or, equivalently, that every repeated transform by $q_1, \cdots, q_m$ (not necessarily in order) of an element of U lies in U_m.

We shall prove this by an induction on the total number of the applied differentiations q_i. The result holds trivially if this number is 0 or 1. Suppose it has been established for all q-transforms of elements of U in which the total number of q_i's is less than s. Let v be an s-tuple transform of an element $u \varepsilon U$. Since the q_i are images by the isomorphism β of the $p_i \varepsilon P$, we have relations $q_j q_i = q_i q_j + \sum_{k=1}^{m} c'_{ijk} q_k$, where the c'_{ijk} are elements of F. By using these relations and the inductive hypothesis it is clear that we can show v to differ from an ordered transform $q^{e_m}{}_m \cdots q^{e_1}{}_1\{u\}$ only by an element in U_m. But the ordered transform belongs to U_m, as is evident from the definition of U_m. Hence $v \varepsilon U_m$.

We have proved our main result:

THEOREM 5.1. *Let A be an arbitrary Lie algebra over the field F of characteristic 0. Then A is the direct sum of two Lie algebras T and H, where T is (0) or semisimple, and H can be represented as follows: There is an isomorphism $\bar{h} \to h$ of H onto a Lie algebra of differentiations in $K\langle x_1, \cdots, x_n\rangle$, where n is the dimension of the maximal solvable ideal of H, and K is a finite algebraic extension field of F. There is a finite dimensional subspace W of $K\langle x_1, \cdots, x_n\rangle$ which contains the x_i and which is mapped into itself by every $\bar{h}$. If $\bar{h}$ is the restriction of h to W, the mapping $h \to \bar{h}$ is therefore an isomorphism of H onto a Lie algebra of linear transformations in W, regarding W as a vector space over F.*

Note. Actually, the decomposition $A = T + H$ is no essential restriction. It is easily seen that we can find an isomorphism of A onto a Lie algebra $\bar{A}$ of differentiations in $K\langle x_1, \cdots, x_n; y_1, \cdots, y_s \rangle$, where s is the dimension of T, and there is still a finite dimensional K-linear subspace which yields a faithful linear representation of A. In fact we can arrange matters so that, with W as above, the space $W + y_1 W + \cdots + y_s W$ has this property.

UNIVERSITY OF ILLINOIS.

BIBLIOGRAPHY.

[1] I. Ado, "Ueber die Darstellung der endlichen kontinuierlichen Gruppen durch lineare Substitutionen," *Bulletin de la Societé Physico-Mathématique de Kazan*, vol. 7 (1934-1935), pp. 3-43.

[2] G. Birkhoff, "Representability of Lie algebras and Lie groups by matrices," *Annals of Mathematics*, vol. 38 (1937), pp. 526-532.

[3] E. Cartan, "Les representations linéaires des groupes de Lie," *Journal de Mathématiques pures et appliquées*, vol. 17 (1938), pp. 1-12.

[4] C. Chevalley, *Theory of Lie Groups*, Princeton, 1946.

[5] C. Chevalley, "Algebraic Lie algebras," *Annals of Mathematics*, vol. 48 (1947), pp. 91-100.

[6] C. Chevalley and S. Eilenberg, "Cohomology theory of Lie groups and Lie algebras," *Transactions of the American Mathematical Society*, vol. 63 (1948), pp. 85-124.

[7] Harish-Chandra, "Faithful representations of Lie algebras," *Annals of Mathematics*, vol. 50 (1949).

[8] N. Jacobson, "Restricted Lie algebras of characteristic p," *Transactions of the American Mathematical Society*, vol. 50 (1941), pp. 15-25.

[9] J. H. C. Whitehead, *Proceedings of the Cambridge Philosophical Society*, vol. 32 (1936), pp. 229-237.

SOME THEOREMS ON ALMOST PERIODIC FUNCTIONS.*

By Raouf Doss.

We shall be concerned here with the following classes of almost periodic functions: the class (B) of Besicovitch, its subclass (Bb) of bounded functions, and the class (B_0) of Bohr functions.

An almost periodic function $f(x) \sim \sum b_{u_\nu} e^{iu_\nu x}$ is said to be of basis $\{\beta_i\}$, where the β_i are linearly independent, if each exponent u_ν is a linear combination with rational coefficents of the β_i.

In Theorem I we prove that every linear functional $U(f)$ defined in the space of functions of class (B) and basis $\{\beta_i\}$ is of the form

$$U(f) = \mathfrak{M}\{f(x)\alpha(x)\}$$

where $\alpha(x)$ is a function of the class (Bb).

In Theorem II we prove that every linear functional $U(f)$ defined in the space of functions of class (B_0) and basis $\{\beta_i\}$ is of the form

$$U(f) = \mathfrak{M}\{f(x)\alpha(x)\}$$

where $\alpha(x)$ is a function summable on every finite interval, such that $\mathfrak{M}\{|\alpha(x)|\} < \infty$ and such that $\mathfrak{M}\{e^{iu_\nu x}\alpha(x)\}$ exists for every linear combination u_ν of the β_i.

Theorem I corresponds to the theorem of Steinhaus on the form of the linear functionals defined in the space of summable functions of a given period.

Theorem II corresponds to the case of Riesz. The introduction of the mean value $\mathfrak{M}$ allowed us to dispense with the use of the Stieltjes integral which appears in the Riesz form.[1]

In Theorem III we complete an investigation made some years ago.[2] We give the necessary and sufficient condition that a trigonometric series $\sum b_{u_\nu} e^{iu_\nu x}$ should be the development of some function of the class (B). Using an expression of Kovanko, the condition is that the Bochner sums associated with the series are equally B-uniformly summable.

---

* Received August 2, 1948.

[1] A complete analogue to the Riesz form is given in S. Bochner, "Additive set functions on groups," *Annals of Mathematics*, vol. 40 (1939), pp. 769-799, th. I.

[2] Raouf Doss, "Contribution to the theory of almost periodic functions," *Annals of Mathematics*, vol. 46 (1945), pp. 196-219 (quoted in the sequel as "Contribution").

The method of proof is the same throughout. It is based on the use of the Bochner sums.

1. We fix first our notations. $f(x)$ being summable on every finite interval, we write

$$\bar{\mathfrak{M}}_x\{f(x)\} = \bar{\mathfrak{M}}\{f(x)\} = \limsup\, (2T)^{-1} \int_{-T}^{T} f(x)\, dx$$

as $T \to \infty$, and $\mathfrak{M}\{f(x)\} = \lim\, (2T)^{-1} \int_{-T}^{T} f(x)\, dx$ if this last limit exists.

For any trigonometric series

$$\Sigma b_{u_n} e^{iu_n x} \tag{1}$$

of basis $\{\beta_i\}$, we define the Bochner sums $\tau_m(x)$ as

$$\tau_m(x) = \sum_{\nu_1=-m!^2}^{\nu_1=m!^2} \cdots \sum_{\nu_m=-m!^2}^{\nu_m=m!^2} (1-|\nu_1|/m!^2) \cdots (1-|\nu_m|/m!^2) b_{u_\nu} e^{iu_\nu x}$$

where $u_\nu = \nu_1(\beta_1/m!) + \cdots + \nu_m(\beta_m/m!)$.

We shall write $\tau_m(x) = \Sigma d^m{}_{u_\nu} b_{u_\nu} e^{iu_\nu x}$. If series (1) is the expansion of some function $f(x) \in (B)$ its Bochner sums will be denoted by $f_m(x)$. In that case $f_m(x) = \mathfrak{M}_t\{f(x+t)K_m(t)\}$ where $K_m(t)$ is the Bochner-Fejér Kernel $K_m(t) = \Sigma d^m{}_{u_\nu} e^{-iu_\nu t}$.

LEMMA I. *Let $\sigma(x)$ be a function of Bohr of basis $\{\beta_i\}$ such that l.u.b. $|\sigma(x)| = M$, where $-\infty < x < \infty$. Then the norm of the functional $T(f) = \mathfrak{M}\{f(x)\sigma(x)\}$ defined in the space (B) is $\|T\| = M$.*

Proof. We have evidently $\|T\| \leq M$ since $|T(f)| \leq M \cdot \mathfrak{M}\{|f(x)|\} = M \cdot \|f\|$. On the other hand, let $\epsilon > 0$ be given and let x_0 be such that $|\sigma(x_0)| > (1-\epsilon)M$. The functions $\sigma_m(x) = \mathfrak{M}_t\{\sigma(x+t)K_m(t)\}$ tend uniformly to $\sigma(x)$ as $m \to \infty$. Hence, for sufficiently large m, $|\sigma_m(x_0)| > (1-\epsilon)M$. But

$$\sigma_m(x_0) = \mathfrak{M}_t\{K_m(t-x_0)\sigma(t)\} = \mathfrak{M}_x\{K_m(x-x_0)\sigma(x)\}.$$

We conclude

$$\begin{aligned}(1-\epsilon)M < |\sigma_m(x_0)| &= |T(K_m(x-x_0))| \leq \|T\| \cdot \|K_m(x-x_0)\| \\ &= \|T\| \cdot \mathfrak{M}_x\{K_m(x-x_0)\} = \|T\| \cdot \mathfrak{M}_x\{K_m(x)\} = \|T\|.\end{aligned}$$

ϵ being arbitrary, this gives $\|T\| = M$.

THEOREM I. *Any linear functional* $U(f)$ *defined in the space of functions of the class* (B) *and of basis* $\{\beta_i\}$ *is of the form* $U(f) = \mathfrak{M}\{f(x)\alpha(x)\}$ *where* $\alpha(x) \varepsilon (Bb)$.

Proof. For any linear combination u_n of the β_i put

$$U(e^{iu_nx}) = a_{u_n} = \mathfrak{M}\{e^{iu_nx}a_{u_n}e^{-iu_nx}\}.$$

The Bochner sum associated with the series

$$\Sigma a_{u_n}e^{-iu_nx} \tag{2}$$

will be denoted $\tau_m(x)$. Since $d^m{}_{-u_n} = d^m{}_{u_n}$ we may write

$$\tau_m(x) = \Sigma d^m{}_{u_\nu}a_{-u_\nu}e^{iu_\nu x} = \Sigma d^m{}_{u_\nu}\, a_{u_\nu}e^{-iu_\nu x}$$

We have, for any $f(x) \sim \Sigma b_{u_\nu}e^{iu_\nu x}$ of the class (B), in virtue of the linearity of U,

$$U(f_m(x)) = \Sigma d^m{}_{u_\nu}b_{u_\nu}a_{u_\nu} = \mathfrak{M}\{f(x)\tau_m(x)\}.$$

If we show that series (2) is the expansion of some function $\alpha(x) \varepsilon (Bb)$ with $|\alpha(x)| \leq A$, then

$$U(f_m(x)) = \mathfrak{M}\{f_m(x)\alpha(x)\}.$$

But

$$|\mathfrak{M}\{[f(x) - f_m(x)]\alpha(x)\}| \leq A \cdot \mathfrak{M}\{|f(x) - f_m(x)|\}.$$

Since $\| f - f_m \| = \mathfrak{M}\{|f(x) - f_m(x)|\}$ tend to 0 with $1/m$, then, by the continuity of U

$$U(f) = \lim U(f_m) = \lim \mathfrak{M}\{f_m(x)\alpha(x)\} = \mathfrak{M}\{f(x)\alpha(x)\}, \text{ where } m \to \infty,$$

and our theorem will be proved.

To prove that series (2) is the expansion of some function $\alpha(x) \varepsilon (Bb)$, with $|\alpha(x)| \leq A$, it is sufficient to show[3] that

$$|\tau_m(x)| \leq A \tag{3}$$

for every m and x. The linear functionals

$$U_m(f) = U(f_m) = \mathfrak{M}\{f(x)\tau_m(x)\}$$

are weakly convergent, since for every $f \varepsilon (B)$ of basis $\{\beta_i\}$, $\lim U_m(f) = U(f)$, where $m \to \infty$. By a well known theorem of Banach and Steinhaus,[4] their norms $\| U_m \| = \text{u.b.} \, | \tau_m(x) |$ are bounded by some constant A. Then (3) is true and the theorem is proved.

---

[3] See "Contribution," th. VI.

2. THEOREM II. *Every linear functional* $U(f)$ *defined in the space of functions of the class* (B_0) *and basis* $\{\beta_i\}$ *is of the form*

$$(1) \qquad U(f) = \mathfrak{M}\{f(x)\alpha(x)\},$$

where $\alpha(x)$ *is a function summable on every finite interval, such that* $\mathfrak{M}\{| \alpha(x) |\} < \infty$ *and such that* $\mathfrak{M}\{e^{iu_\nu x}\alpha(x)\}$ *exists for every linear combination* u_ν *of the* β_i. *Conversely, for any function* $\alpha(x)$ *with the prescribed conditions,* $\mathfrak{M}\{f(x)\alpha(x)\}$ *exists and is a linear functional in the space* (B_0) *of basis* $\{\beta_i\}$.

Proof. Put, as before $U(e^{iu_\nu x}) = a_{u_\nu} = \mathfrak{M}\{e^{iu_\nu x} a_{u_\nu} e^{-iu_\nu x}\}$ and let $\tau_m(x)$ be the Bochner sums associated with the series $\Sigma a_{u_\nu} e^{-iu_\nu x}$.

Then, for every $f(x) \, \varepsilon \, (B_0)$, of basis $\{\beta_i\}$,

$$U(f) = \lim \mathfrak{M}\{f(x)\tau_m(x)\}, \text{ where } m \to \infty,$$

is finite. We conclude[5] that there exists a constant M such that

$$\mathfrak{M}\{| \tau_m(x) |\} \leq M \quad (\text{for every } m).$$

We know that for every $\tau(x) \, \varepsilon \, (B_0)$ the integral $(1/L) \int_a^{a+L} \tau(x)\,dx$ tends, uniformly in a, to $\mathfrak{M}\{\tau(x)\}$ as $L \to \infty$. Let $\{\epsilon_n\}$ be a decreasing sequence of positive numbers tending to 0. Since $| \tau_m(x) | \, \varepsilon \, (B_0)$, we can determine a sequence $\{L_m\}$ such that for $L \geq L_m$ and every a

$$(2) \qquad 1/L \int_a^{a+L} | \tau_m(x) | \, dx < M + \epsilon_m.$$

Similarly, since $e^{iu_\nu x}\tau_m(x) \, \varepsilon \, (B_0)$, we can determine a sequence $L^\nu{}_1, L^\nu{}_2, \cdots, L^\nu{}_m, \cdots$ such that

$$| 1/L \int_a^{a+L} e^{iu_\nu x}\tau_m(x)\,dx - \mathfrak{M}\{e^{iu_\nu x}\tau_m(x)\} | < \epsilon_m,$$

i. e.

$$(3) \qquad | 1/L \int_a^{a+L} e^{iu_\nu x}\tau_m(x) - d^m{}_{u_\nu} a_{u_\nu} | < \epsilon_m.$$

for $L \geq L^\nu{}_m$ and every a.

---

[4] See, for example, S. Banach, *Théorie des Opérations Linéaires*, Varsovie, 1931, p. 80, th. 5.

[5] See "Contribution," Lemma, p. 210.

Let M_m be the upper bound of $|\tau_m(x)|$. We shall suppose that the following relations, bearing on the diagonal elements $L^m{}_m$ and requiring these elements to be sufficiently large, are satisfied:

$$(4)\quad L_m/L^m{}_m \leqq 1;\quad (5)\quad L^\nu{}_m/L^m{}_m \leqq 1, \nu < m;$$

$$(6)\quad L^\nu{}_m/L^{m-1}{}_{m-1} \leq \epsilon_{m-1},\ \nu < m-1;$$

$$(7)\quad L^\nu{}_m M_m/L^{m-1}{}_{m-1} \leq \epsilon_{m-1},\ \nu < m-1;\quad (8)\quad L_m M_m/L^{m-1}{}_{m-1} \leq \epsilon_{m-1}.$$

Starting from the origin, we shall put, on the right, contiguous intervals $(L^1{}_1), (L^2{}_2), \cdots, (L^n{}_n), \cdots$ of lengths $L^1{}_1, L^2{}_2, \cdots, L^n{}_n, \cdots$ and on the left, contiguous intervals $(-L^1{}_1), (-L^2{}_2), \cdots, (-L^n{}_n), \cdots$ of the same lengths. Let $\alpha(x)$ be a function equal to $\tau_m(x)$ if $x \varepsilon (L^m{}_m)$ or if $x \varepsilon (-L^m{}_m)$.

We shall show that $\mathfrak{M}\{|\alpha(x)|\} \leq M + 2\epsilon_1 = M'$.

In fact, let $(-T, T)$ be any interval and suppose that the right end of this interval is in the interval $(L^n{}_n)$, covering a part (l_n) of it, of length l_n. We have

$$\int_0^T |\alpha(x)|\,dx = \int_{(L^1{}_1)} |\tau_1(x)|\,dx + \cdots + \int_{(L^{n-1}{}_{n-1})} |\tau_{n-1}(x)|\,dx$$
$$+ \int_{(l_n)} |\tau_n(x)|\,dx.$$

and we have an analogous relation for $\int_{-T}^0 |\alpha(x)|\,dx$.

We consider two cases:

First case. $l_n \geq L_n$. Then by (2) and (4)

$$(9)\qquad \int_0^T |\alpha(x)|\,dx \leq L^1{}_1(M+\epsilon_1) + \cdots + L^{n-1}{}_{n-1}(M+\epsilon_{n-1})$$
$$+ l_n(M+\epsilon_n) \leq T(M+\epsilon_1).$$

Second case. $l_n < L_n$. Then, by (8)

$$(10)\qquad \int_0^T |\alpha(x)|\,dx \leq [L^1{}_1 + \cdots + L^{n-1}{}_{n-1}](M+\epsilon_1) + l_n M_n$$
$$\leq T(M+\epsilon_1) + \epsilon_{n-1} L^{n-1}{}_{n-1} \leq T(M+\epsilon_1+\epsilon_{n-1}).$$

The two relations (9) and (10), together with the analogous relation for $\int_{-T}^0 |\alpha(x)|\,dx$ show that $\mathfrak{M}\{|\alpha(x)|\} \leq M + 2\epsilon_1 = M'$.

We shall now show that, for every u_ν, $\mathfrak{M}\{e^{iu_\nu x}\alpha(x)\}$ exists and is equal to a_{u_ν}.

With the preceding notations, we have, for $n > \nu + 1$, i. e., for sufficiently large T,

$$\int_0^T e^{iu_\nu x}\alpha(x)\,dx = \int_{(L^1{}_1)} e^{iu_\nu x}\tau_1(x)\,dx + \cdots + \int_{(L^{\nu-1}{}_{\nu-1})} e^{iu_\nu x}\tau_{\nu-1}(x)\,dx$$
$$+ \int_{(L^\nu{}_\nu)} e^{iu_\nu x}\tau_\nu(x)\,dx + \cdots + \int_{(L^{n-1}{}_{n-1})} e^{iu_\nu x}\tau_{n-1}(x)\,dx + \int_{(l_n)} e^{iu_\nu x}\tau_n(x)\,dx.$$

We consider again two cases:

First case. $l_n \geqq L^\nu{}_n$. We have, by (3) and (5)

$$\int_0^T e^{iu_\nu x}\alpha(x)\,dx = \int_{(L^1{}_1)} e^{iu_\nu x}\tau_1(x)\,dx + \cdots + \int_{(L^{\nu-1}{}_{\nu-1})} e^{iu_\nu x}\tau_{\nu-1}(x)\,dx$$
$$+ L^\nu{}_\nu(d^\nu{}_{u_\nu}a_{u_\nu} + \theta_\nu\epsilon_\nu) + \cdots + L^{n-1}{}_{n-1}(d^{n-1}{}_{u_\nu}a_{u_\nu} + \theta_{n-1}\epsilon_{n-1})$$
$$+ l_n(d^n{}_{u_\nu}a_{u_\nu} + \theta_n\epsilon_n), \tag{11}$$

where $\theta_\nu, \theta_{\nu+1}, \cdots, \theta_n$ have moduli $\leqq 1$.

Second case. $l_n < L^\nu{}_n$. Then, by (7)

$$\left| \int_{(l_n)} e^{iu_\nu x}\tau_n(x)\,dx \right| \leqq l_n M_n < L^\nu{}_n M_n \leqq L^{n-1}{}_{n-1}\epsilon_{n-1}.$$

Hence

$$\int_0^T e^{iu_\nu x}\alpha(x)\,dx = \int_{(L^1{}_1)} e^{iu_\nu x}\tau_1(x)\,dx + \cdots + \int_{(L^{\nu-1}{}_{\nu-1})} e^{iu_\nu x}\tau_{\nu-1}(x)\,dx$$
$$+ L^\nu{}_\nu(d^\nu{}_{u_\nu}a_{u_\nu} + \theta_\nu\epsilon_\nu) + \cdots + L^{n-1}{}_{n-1}(d^{n-1}{}_{u_\nu}a_{u_\nu} + \theta_{n-1}\epsilon_{n-1}), \tag{12}$$

where $\theta_\nu, \theta_{\nu+1}, \cdots, \theta_{n-2}$ have moduli $\leqq 1$ and θ_{n-1} a modulus $\leqq 2$.

Divide each of equations (11) and (12) by T and put

$$\lambda_m{}^T = L^m{}_m/T \ (\nu \leqq m < n); \qquad \lambda_m{}^T = l_n/T \ (m = n).$$

As $m \to \infty$, we have $\lim d^m{}_{u_\nu} = 1$, i. e., $\lim\,(d^m{}_{u_\nu}a_{u_\nu} + \theta_m\epsilon_m) = a_{u_\nu}$, and this limit is uniform in T, since the $\theta_m = \theta_m(T)$ are all in absolute value less than 2.

If we show that the $\lambda_m{}^T$ are the coefficients of a regular process of summation of Toeplitz, then

$$\lim 1/T \int_0^T e^{iu_\nu x}\alpha(x)\,dx = a_{u_\nu}, \text{ where } T \to \infty. \tag{13}$$

We have to verify two conditions:

1) $\lim \lambda_m{}^T = 0$, where $T \to \infty$ for every m. This is evident.

2) In the first case we have

$$\lim \sum_{m=\nu}^{n} \lambda_m{}^T = \lim \{T - (L^1{}_1 + \cdots + L^{\nu-1}{}_{\nu-1})\}/T = 1, \text{ where } T \to \infty.$$

In the second case we have

$$\sum_{m=\nu}^{n-1} \lambda_m{}^T = \{\mathrm{T} - (L^1{}_1 + \cdots + L^{\nu-1}{}_{\nu-1})\}/T - l_n/T;$$

but, by (6), $l_n/T < L^\nu{}_n/T < L^\nu{}_n/L^{n-1}{}_{n-1} \leq \epsilon_{n-1}$, so that also

$$\lim \sum_{m=\nu}^{n-1} \lambda_m{}^T = 1, \text{ where } T \to \infty.$$

Relation (13) is thus satisfied; but we have an analogous relation for the integral from $-T$ to 0. Finally

$$\mathfrak{M}\{e^{iu_\nu x}\alpha(x)\} = a_{u_\nu} = U(e^{iu_\nu x}). \tag{14}$$

It remains to be verified that for every $f(x) \varepsilon (B_0)$ of basis $\{\beta_i\}$, $\mathfrak{M}\{f(x)\alpha(x)\}$ exists and is equal to $U(f)$.

For an arbitrary $\epsilon > 0$ we can find a large m such that

$$| f(x) - f_m(x) | < \epsilon/(M' + \epsilon), \qquad | U(f) - U(f_m) | < \epsilon. \tag{15}$$

We have, by (14), $U(f_m) = \mathfrak{M}\{f_m(x)\alpha(x)\}$. Take T_0 such that for $T > T_0$, $(2T)^{-1} \int_{-T}^{T} | \alpha(x) | \, dx < M' + \epsilon$ and

$$| U(f_m) - (2T)^{-1} \int_{-T}^{T} f_m(x)\alpha(x)\,dx \,| < \epsilon. \tag{16}$$

Then

$$| (2T)^{-1} \int_{-T}^{T} [f_m(x) - f(x)]\alpha(x)\,dx \,| \leq (\epsilon/M' + \epsilon) \cdot (M' + \epsilon) = \epsilon. \tag{17}$$

The three relations (15), (16), (17) show that, for $T > T_0$,

$$| U(f) - (2T)^{-1} \int_{-T}^{T} f(x)\alpha(x)\,dx \,| < 3\epsilon.$$

Whence $U(f) = \mathfrak{M}\{f(x)\alpha(x)\}$.

The first part of the theorem is now proved. The second part presents no difficulty.

3. LEMMA 2. *Let* $\alpha^n(x) \sim \Sigma a^n{}_{u_\nu} e^{iu_\nu x}$ *be a sequence of functions of the* *class* (Bb) *of basis* $\{\beta_i\}$ *bounded by the same constant* A *and such that* $\lim a^n_{u_\nu} = a_{u_\nu}$, where $n \to \infty$. *Then*

$$\Sigma a_{u_\nu} e^{iu_\nu x} \tag{1}$$

is the expansion of some function $\alpha(x) \epsilon (Bb)$ *and bounded by* A. *Moreover* $\mathfrak{M}\{|\alpha(x)|\} \leq \text{l. u. b.}\, \mathfrak{M}\{|\alpha^n(x)|\} = B$ *and*

$$\mathfrak{M}\{f(x)\alpha(x)\} = \lim \mathfrak{M}\{f(x)\alpha^n(x)\}$$

where $n \to \infty$, *for every* $f(x) \varepsilon (Bb)$ *of basis* $\{\beta_i\}$.

Proof. Let $\epsilon > 0$ be given and let $\alpha_m(x)$ be the Bochner sums associated with (1). We have

$$|\alpha_m(x) - \alpha^n{}_m(x)| \leq \epsilon \qquad \text{(for } m \text{ fixed and } n > n_0). \tag{2}$$

Hence

$$|\alpha_m(x)| \leq |\alpha^n{}_m(x)| + \epsilon \leq A + \epsilon,$$

which proves the first part of the lemma. If now m is chosen such that $\mathfrak{M}\{|\alpha(x) - \alpha_m(x)|\} \leq \epsilon$, then

$$\mathfrak{M}\{|\alpha(x)|\} \leq \mathfrak{M}\{|\alpha^n{}_m(x)|\} + 2\epsilon \leq B + 2\epsilon. \tag{3}$$

Finally, if $|f(x)| < C$ and if m is chosen such that

$$\mathfrak{M}\{|f(x) - f_m(x)|\} < \epsilon, \tag{4}$$

then, for $n > n_0$, by (4) and (2)

$$|\mathfrak{M}\{f[\alpha - \alpha^n]\}| \leq |\mathfrak{M}\{[f - f_m][\alpha - \alpha^n]\}| + |\mathfrak{M}\{f_m[\alpha - \alpha^n]\}|$$
$$\leq 2A \cdot \mathfrak{M}\{|f - f_m|\} + |\mathfrak{M}\{f[\alpha_m - \alpha^n{}_m]\}| \leq 2A \cdot \epsilon + C \cdot \epsilon.$$

This completes the proof.

COROLLARY. *Let* $\alpha^n(x)$ *be functions of the class* (Bb) *of basis* $\{\beta_i\}$, *bounded by the same constant* A. *Then we can extract a partial sequence* $\{\alpha^{n_k}(x)\}$ *and find a function* $\alpha(x) \varepsilon (Bb)$ *bounded by* A, *such that*

$$\mathfrak{M}\{f(x)\alpha(x)\} = \lim \mathfrak{M}\{f(x)\alpha^{n_k}(x)\}, \text{ where } k \to \infty,$$

for every $f(x) \varepsilon (Bb)$ *of basis* $\{\beta_i\}$ *and we have*

$$\mathfrak{M}\{|\alpha(x)|\} \leq \text{l. u. b.}\, \mathfrak{M}\{|\alpha^{n_k}(x)|\}.$$

The partial sequence is obtained by the diagonal process. Note also that by again extracting a subsequence, "upper bound" can be replaced by "limit" and we have

$$\mathfrak{M}\{|\alpha(x)|\} \leq \lim \mathfrak{M}\{|\alpha^{n_k}(x)|\}, \text{ where } k \to \infty.$$

Notation.[6] Given a measurable set E, a summable function $f(x)$ and any interval $(-T, T)$, we mean by $E(-T, T)$ the common part of E and $(-T, T)$ and we write

$$\bar{\delta}E = \limsup mesE(-T, T)/2T, \text{ where } T \to \infty,$$

and

$$\bar{\mathfrak{M}}^E{}_x\{f(x)\} = \limsup\ (2T)^{-1} \int_{E(-T,T)} f(x)\,dx, \text{ where } T \to \infty.$$

THEOREM III. *In order that the series* $\Sigma a_{u_\nu} e^{iu_\nu x}$ *be the expansion of some function* $\alpha(x)\ \varepsilon\ (B)$, *it is necessary and sufficient that the Bochner sums* $\sigma_m(x)$ *attached to the series satisfy the following condition: To every* $\epsilon > 0$ *there corresponds an* $\eta > 0$ *such that* $\bar{\mathfrak{M}}^E{}_x\{|\sigma_m(x)|\} \leq \epsilon$ *for every* m *and every* E *for which* $\bar{\delta}E \leq \eta$.

Proof. Necessity. Let the given series be the expansion of $\alpha(x)\ \varepsilon\ (B)$ and let $\epsilon > 0$ be given. We know that there exists an m_0 such that for every $m \geq m_0$, $\mathfrak{M}\{|\alpha(x) - \sigma_m(x)|\} \leq \epsilon/4$. Whence $\mathfrak{M}\{|\sigma_m(x) - \sigma_{m_0}(x)|\} \leq \epsilon/2$, $(m \geq m_0)$. The polynomial $\sigma_{m_0}(x)$ being bounded, say by the constant A, choose $\eta' = \epsilon/2A$. Hence, for any set E for which $\bar{\delta}E \leq \eta'$,

$$\bar{\mathfrak{M}}^E{}_x\{|\sigma_{m_0}(x)|\} \leq A \cdot \bar{\delta}E \leq A \cdot \epsilon/2A = \epsilon/2,$$

so that, for $m \geq m_0$

$$\bar{\mathfrak{M}}^E{}_x\{|\sigma_m(x)|\} \leq \bar{\mathfrak{M}}_x\{|\sigma_m(x) - \sigma_{m_0}(x)|\} + \bar{\mathfrak{M}}^E{}_x\{|\sigma_{m_0}(x)|\} \leq \epsilon.$$

The number η' therefore verifies the condition of the theorem for $m \geq m_0$. But there is only a finite number of $\sigma_m(x)$ with $m < m_0$ and each of them is bounded. Let A' be an upper bound for $|\sigma_m(x)|$ for $m < m_0$. The number $\eta = \min\{\eta', \epsilon/A'\}$ verifies the condition of the theorem.

Sufficiency. By the condition of the theorem there exists a constant B such that

$$\mathfrak{M}\{|\sigma_m(x)|\} \leq B \qquad \text{(for every } m).$$

---

[6] Kovanko, "Sur la structure des fonctions presque périodiques généralisées," *Recueil Mathématique*, Moscow, vol. 42 (1935), pp. 3-10.

Let n be fixed. Put

$$\sigma^n{}_m(x) = \sigma_m(x) \text{ if } |\sigma_m(x)| \leq n; \quad \sigma^n{}_m(x) = n \cdot \sigma_m(x)/|\sigma_m(x)| \quad \text{if } |\sigma_m(x)| > n.$$

Then, $\sigma^n{}_m(x) \varepsilon (B)$.[7]

The functions $\sigma^n{}_m(x)$ form an enumerable set, so that we may suppose that, together with the given series $\Sigma a_{u_\nu} e^{iu_\nu x}$, they belong to the same basis $\{\beta_i\}$. The sequence $\{\sigma^n{}_m(x)\}_m$ being bounded, there exists a partial sequence $\{\sigma^n{}_{m_k}(x)\}_k$ and a bounded function $\sigma^n(x) \varepsilon (Bb)$ such that

$$\mathfrak{M}\{|\sigma^n(x)|\} \leq \lim \mathfrak{M}\{|\sigma^n{}_{m_k}(x)|\}$$

and

$$\mathfrak{M}\{f(x)\sigma^n(x)\} = \lim \mathfrak{M}\{f(x)\sigma^n{}_{m_k}(x)\}, \text{ where } k \rightarrow \infty, \tag{5}$$

for every $f(x) \varepsilon (Bb)$ of basis $\{\beta_i\}$.

Making use of the diagonal process we might suppose that the same subsequence $\{m_k\}$ is valid for all n and even for all pairs (p, q) since these pairs form an enumerable set, so that we shall write

$$\mathfrak{M}\{|\sigma^p(x) - \sigma^q(x)|\} \leq \lim \mathfrak{M}\{|\sigma^p{}_{m_k}(x) - \sigma^q{}_{m_k}(x)|\}, \text{ where } k \rightarrow \infty. \tag{6}$$

Let now E_{m,n_0} be the set of points for which $|\sigma_m(x)| \geq n_0$ and let n_0 be such that $B/n_0 < \eta$. We have $n_0 \cdot \bar{\delta} E_{m,n_0} \leq \bar{\mathfrak{M}}^{E_{m,n_0}}{}_x\{|\sigma_m(x)|\} \leq B$ i.e., $\bar{\delta} E_{m,n_0} \leq B/n_0 < \eta$. Hence

$$\bar{\mathfrak{M}}^{E_{m,n_0}}{}_x\{|\sigma_m(x)|\} \leq \epsilon; \ \bar{\mathfrak{M}}^{E_{m,n_0}}{}_x\{|\sigma^n{}_m(x)|\} \leqq \epsilon. \tag{7}$$

We conclude, by (6), for $p, q > n_0$

$$\begin{aligned} \mathfrak{M}\{|\sigma^p(x) - \sigma^q(x)|\} &\leq \text{l. u. b.}\, \mathfrak{M}\{|\sigma^p{}_{m_k}(x) - \sigma^q{}_{m_k}(x)|\} \\ &\leq \text{l. u. b.}\, \bar{\mathfrak{M}}^{E_{m_k,n_0}}{}_x\{|\sigma^p{}_{m_k}(x) - \sigma^q{}_{m_k}(x)|\} \leq 2\epsilon, \end{aligned}$$

since for $x \notin E_{m_k,n_0}$ we have $|\sigma_{m_k}(x)| < n_0$, i. e. $\sigma^p{}_{m_k}(x) = \sigma_{m_k}(x) = \sigma^q{}_{m_k}(x)$. The sequence $\{\sigma^p(x)\}$ is therefore a Cauchy sequence in the space (B). (B) being complete, $\sigma^p(x)$ converges to some function $\alpha(x) \varepsilon (B)$.

It remains to be shown that the expansion of $\alpha(x)$ is the given series. We can find n_1 such that for $n > n_1$

$$\mathfrak{M}\{|\alpha(x) - \sigma^n(x)|\} < \epsilon. \tag{8}$$

---

[7] See, for example, Kovanko, *loc. cit.*, p. 6.

By (7), for $n > n_0$ and every m_k,

$$(9) \qquad \mathfrak{M}\{|\sigma^n{}_{m_k}(x) - \sigma_{m_k}(x)|\} = \mathfrak{M}^{E_{m_k,n_0}}{}_x\{|\sigma^n{}_{m_k}(x) - \sigma_{m_k}(x)|\} \leq 2\epsilon.$$

By (5), for fixed $n > n_0, n_1$ we can find m_{k_0} such that for $m_k > m_{k_0}$

$$(10) \qquad |\mathfrak{M}\{e^{-iu_\nu x}[\sigma^n(x) - \sigma^n{}_{m_k}(x)]\}| < \epsilon.$$

The three relations (8), (10) and (9) show that for $m_k > m_{k_0}$

$$\begin{aligned} |\mathfrak{M}\{e^{-iu_\nu x}[\alpha(x) - \sigma_{m_k}(x)]\}| &\leq |\mathfrak{M}\{e^{-iu_\nu x}[\alpha(x) - \sigma^n(x)]\}| \\ &+ |\mathfrak{M}\{e^{-iu_\nu x}[\sigma^n(x) - \sigma^n{}_{m_k}(x)]\}| + |\mathfrak{M}\{e^{-iu_\nu x}[\sigma^n{}_{m_k}(x) - \sigma_{m_k}(x)]\}| \\ &\leq \epsilon + \epsilon + 2\epsilon. \end{aligned}$$

Therefore

$$\mathfrak{M}\{e^{-iu_\nu x}\alpha(x)\} = \lim \mathfrak{M}\{\{e^{-iu_\nu x}\sigma_{m_k}(x)\} = \lim d^{m_k}{}_{u_\nu} a_{u_\nu} = a_{u_\nu},$$

where $k \to \infty$.

This completes the proof of the theorem.

Remark 1. We conclude from the preceding theorem that if

$$(11) \qquad f(x) \sim \Sigma b_{u_\nu} e^{iu_\nu x}$$

is an arbitrary function of the class (B) and if $\{\gamma_i\}$ is an arbitrary sequence of linearly independent numbers, then the subseries (Γ) of (11), corresponding to the u_ν which are linear combinations of the γ_i, is the expansion of some function $f^{(\gamma)}(x) \,\varepsilon\, (B)$.[8] For if $H_m(t)$ is the Bochner-Fejér kernel corresponding to $\{\gamma_i\}$ then the sequence $\sigma^{(\gamma)}{}_m(x) = \mathfrak{M}_t\{f(x+t)H_m(t)\}$ is associated to the subseries (Γ) and satisfies the condition of Theorem III. $f^{(\gamma)}(x)$ may be called the restriction of $f(x)$ to the basis $\{\gamma_i\}$.

Remark 2. Theorem I may now be generalized in the sense if $U(f)$ is defined for *all* functions of class (B) then, still $U(f) = \mathfrak{M}\{f(x)\alpha(x)\}$ where $\alpha(x) \,\varepsilon\, (Bb)$.

In fact, put $U(e^{iux}) = a_u$. Then the set of values of u for which $a_u \neq 0$ is at most enumerable. For otherwise, there exists an $a > 0$ such that for a non-enumerable infinity of u: $|a_u| > a$. We can select from these u a sequence of linearly independent numbers $\gamma_1, \gamma_2, \cdots, \gamma_i, \cdots$ such that $|a_{\gamma_i}| > a \quad (i = 1, 2, \cdots)$. Also $|a_{\gamma_i}| \leq \| U \|$. The function $f(x) \sim \Sigma (1/n) \bar{a}_{\gamma_n} e^{i\gamma_n x}$ belongs to the class (B^2) of Besicovitch and hence to the class (B). If $f_m(x)$ are its Bochner sums we have

---

[8] The corresponding statement for functions of the class (B_0) is due to S. Bochner: "Beiträge zur Theorie der fastperiodische Funktionen," *Mathematische Annalen*, vol. 96 (1926), pp. 119-147.

$$f_m(x) = \sum_{n=1}^{m} (1 - (1/m!)) (1/n) \bar{a}_{\gamma_n} e^{i\gamma_n x}$$

and

$$U(f) = \lim U(f_m) = \lim \sum_{n=1}^{m} (1 - (1/m!)) (1/n) | a_{\gamma_n} |^2 = \infty,$$

where $m \to \infty$

which is impossible.

Let then $\beta_1, \beta_2, \cdots, \beta_n, \cdots$ be a basis for the enumerable set of u's for which $a_u \neq 0$. The linear combinations with rational coefficients of the β_i will be denoted $u_1, u_2, \cdots, u_\nu, \cdots$ so that $a_u = 0$ if u is not some u_ν. Then, as in Theorem I $\Sigma a_{u_\nu} e^{-iu_\nu x}$ is the Fourier series of some function $\alpha(x) \varepsilon (Bb)$ and for any $f(x) \varepsilon (B)$ of basis $\{\beta_i\} : U(f) = \mathfrak{M}\{f(x)\alpha(x)\}$. If $f(x)$ has no exponent equal to some u_ν then $U(f) = 0$. Also in that case $\mathfrak{M}\{f(x)\alpha(x)\} = 0$. Finally, if $f(x)$ is any function of class (B), then putting $f(x) = f^{(\beta)}(x) + f(x) - f^{(\beta)}(x)$ we have

$$U(f) = U(f^{(\beta)}) + U(f - f^{(\beta)}) = U(f^{(\beta)}) = \mathfrak{M}\{f^{(\beta)}(x)\alpha(x)\}$$
$$= \mathfrak{M}\{f^{(\beta)}(x)\alpha(x)\} + \mathfrak{M}\{[f(x) - f^{(\beta)}(x)]\alpha(x)\} = \mathfrak{M}\{f(x)\alpha(x)\}.$$

FAROUK I UNIVERSITY,
ALEXANDRIA, EGYPT.

APPLICATION OF A RADICAL OF BROWN AND McCOY TO NON-ASSOCIATIVE RINGS.*

By Malcolm F. Smiley.

---

1. Introduction. Our first purpose in this paper is to point out that the theory of "radicals" of an associative ring as given by Brown and McCoy [1] applies without change to non-associative rings. We then examine the relation of a particular radical [2] to those defined by A. A. Albert [3] for non-associative algebras and by Max Zorn [4] for hypercomplex alternative rings. We find that the radical [2] of Brown and McCoy is the same as that of Albert for a non-associative algebra *which has a unit element* and also that the radical of Brown and McCoy is the same as Zorn's for a hypercomplex alternative ring.

Our résumé of the theory of Brown and McCoy is preceded by a brief outline of the fundamentals of the theory of non-associative rings and of subdirect sums of such rings. The results of the theory of Brown and McCoy for non-associative rings are then stated without further proof. We conclude our paper with a discussion of the relations between the radical we shall use and those of Albert and of Zorn.

We are indebted to Professor McCoy for directing our attention to his theory and for a stimulating correspondence during the preparation of this paper.

2. Fundamental properties of non-associative rings. In this section we shall briefly outline the facts concerning non-associative rings which we shall need in our ensuing development. All of these facts are well known and we include their statement only for the sake of completeness.

A non-associative ring (*naring*) R is an algebraic system with two single-valued operations $a + b$ and ab defined and in R for every $a, b \varepsilon R$ and such that the system $(R, +)$ is an abelian group and the distributive laws $a(b + c) = ab + ac$ and $(a + b)c = ac + bc$ hold for every $a, b, c \varepsilon R$. It is easy to prove that $0a = a0 = 0, -(-a) = a, (-a)b = a(-b) = -(ab)$,

---

* Received September 3, 1948.

[1] Brown and McCoy [1]. See also McCoy [1].

[2] The F_1-radical in the notation of Brown and McCoy [1].

[3] Albert [1].

[4] Max Zorn [2].

and $(-a)(-b) = ab$ for every a, $b \varepsilon R$, where 0 and $-a$ denote the unit and the inverse of a, respectively, in the abelian group $(R, +)$.

If R is a naring, then a subset M of R is called an *ideal of* R in case for every a, $b \varepsilon M$ and every $x \varepsilon R$ we have $a - b$, ax, and $xa \varepsilon M$. Then M is empty or M is a subgroup of the additive group $(R, +)$ of R and the cosets $\bar{a} = a + M$ constitute a naring if we define $\bar{a} + \bar{b} = \overline{a + b}$ and $\bar{a}\bar{b} = \overline{ab}$. We call the naring of cosets $\bar{a}$ the *difference naring of* R *and* M and we denote this naring by $R - M$. The mapping $a \rightarrow aH = a + M$ is a homomorphism (the *natural* homomorphism) of R onto $R - M$. If R and $\bar{R}$ are narings and $a \rightarrow aT \varepsilon \bar{R}$ is a homomorphism of R onto $\bar{R}$, then the *kernel of* T is the set M of elements $a \varepsilon R$ for which $aT = 0$. Then M is an ideal of R and $R - M \cong \bar{R}$ via $a + M \rightarrow aT$. Each ideal N of $R - M$ gives rise to an ideal $N_0 = [x; x + M \varepsilon N]$ of R and then $N \cong N_0 - M$. We shall also need the fact that if M and N are ideals of R for which $N \geqq M$, then $(R - M) - (N - M) \cong (R - N)$.

If R is a naring, we shall denote the ideal of R generated by an element $a \varepsilon R$ by $I(a)$. More generally, if S is a subset of R we shall denote the ideal of R generated by S by $I(S)$. It is clear that $I(S)$ consists of the set of elements of R of the form $\Sigma s_i + \Sigma s_j \bar{U}_j$, where the sums are *finite*, s_i, $s_j \varepsilon S$, and where each U_j is the product of a finite number of right and left multiplications: $x \rightarrow xR_a = xa$ and $x \rightarrow xL_a = ax$ of R. If $a \rightarrow \bar{a} \varepsilon \bar{R}$ is a homomorphism of R onto a naring $\bar{R}$, it is easy to see that $I(\bar{S}) = \overline{I(S)}$.

The ideals of a naring R form a modular lattice when partially ordered by set inclusion. We shall denote the join of two ideals M and N of R by (M, N), while $M \cap N$, as usual, denotes set-theoretic intersection. If M and N are nonvoid ideals of R, then[5] $(M, N) - M \cong N - (M \cap N.)$.

3. Subdirect sums of non-associative rings. For the sake of clarity we shall elaborate in this section the basic theory of subdirect sums as applied to the systems we are considering.

If R_α $(\alpha \varepsilon \Omega)$ are narings, then the totality of functions $(a_\alpha; \alpha \varepsilon \Omega)$ with $a_\alpha \varepsilon R_\alpha$ constitutes a naring S called the *full* direct sum of the narings R_α. A subnaring T of S is called a *subdirect sum* of the narings R_α in case for each $\alpha \varepsilon \Omega$ the homomorphism $H_\alpha : a \rightarrow aH_\alpha = a_\alpha$ satisfies $(T)H_\alpha = R_\alpha$.

LEMMA. *A naring* R *is isomorphic to a subdirect sum* T *of narings* $R_\alpha(\alpha \varepsilon \Omega)$ *if and only if for each* $\alpha \varepsilon \Omega$, R *contains an ideal* M_α *such that* $R - M_\alpha \cong R_\alpha$ *and* $\Pi M_\alpha = 0$.

---

[5] Garrett Birkhoff [1], pp. 47-48.

Proof. Let R be a naring which is isomorphic to a subdirect sum T of narings $R_a(\alpha \varepsilon \Omega)$ via $a \rightarrow aH \varepsilon T$. The kernel M_a of HH_a is an ideal of R and $R - M_a \cong R_a$ since $(R)HH_a = (T)H_a = R_a$. To see that $\Pi M_a = 0$, let $a \varepsilon \Pi M_a$, then $aHH_a = 0$ for every $\alpha \varepsilon \Omega$ so that $(aH)_a = 0$, *i. e.*, $aH = 0$, $a = 0$.

Conversely, let $M_a(\alpha \varepsilon \Omega)$ be a set of ideals of R which satisfy the stated requirements. Let h_a denote the natural homomorphism of R onto $R_a \equiv R - M_a$. Then $S \equiv [(\bar{a}_a ; \alpha \varepsilon \Omega) ; \bar{a}_a \varepsilon R_a]$ is the full direct sum of the narings R_a since $(T)H_a = (R)h_a = R_a$. If $a \varepsilon R$, define $aH = (ah_a ; \alpha \varepsilon \Omega) \varepsilon T$. Then $a \rightarrow aH$ is an isomorphism of R onto T, since if $aH = bH$, then $(a - b)h_a = 0$ for every $\alpha \varepsilon \Omega$ so that $(a - b) \varepsilon \Pi M_a = 0$ and $a = b$. The proof is complete.

A *subdirectly irreducible* naring R is one which is isomorphic to a subdirect sum T of narings $R_a(\alpha \varepsilon \Omega)$ *only if* H_a is an *isomorphism* for some $\alpha \varepsilon \Omega$. Thus R is subdirectly irreducible if and only if the intersection of all nonzero ideals of R is itself a nonzero ideal J of R. For, if this is true, and if R is isomorphic to a subdirect sum of narings $R_a(\alpha \varepsilon \Omega)$, then, $\Pi M_a = 0$, we have $M\alpha = 0$ for some $\alpha \varepsilon \Omega$ and $R \cong R_a$. On the other hand, if $[M_a ; \alpha \varepsilon \Omega]$ is the totality of nonzero ideals of R, then $\Pi M_a = 0$ would imply that R is isomorphic to a subdirect sum of narings $R_a \cong R - M_a(\alpha \varepsilon \Omega)$. But $R_a \cong R$ is possible for no $\alpha \varepsilon \Omega$ since no M_a is zero. Thus R is not subdirectly irreducible.

4. The F-radical of a non-associative ring. In this and the following section we shall restate the theory of Brown and McCoy as it applies to non-associative rings. We shall omit proofs completely since the proof given by Brown and McCoy are not only valid for non-associative rings but should also be easy for the reader to follow in this case in view of the preparatory material of Sections 2 and 3.

We assume that $a \rightarrow F(a)$ is a mapping defined in each naring[6] R to the set of ideals of R which is such that if $a \rightarrow \bar{a} \varepsilon \bar{R}$ is a homomorphism of R onto a naring $\bar{R}$, then $F(\bar{a}) = \overline{F(a)}$. We define the F-radical $N(R, F)$ of R as the set of elements $b \varepsilon R$ such that if $a \varepsilon I(b)$, then $a \varepsilon F(a)$. If $R = N(R, F)$, we call R an *F-radical naring.*

---

[6] It is possible to phrase our definitions and theorems so as to avoid the meta-mathematical difficulty of the "class of all rings." We have found this awkward, however, and prefer the present formulation.

Theorem 1. *The F-radical* $N(R, F)$ *of a naring* R *is the intersection of all ideals* M *of* R *for which* $R - M$ *is subdirectly irreducible and* $N(R - M, F) = 0$.

Corollary 1. *The F-radical of a naring* R *is an ideal of* R.

Corollary 2. *A naring* R *is an F-radical naring if and only if* R *itself is the only ideal* M *of* R *for which* $R - M$ *is subdirectly irreducible and* $N(R - M, F) = 0$.

Theorem 2. *If* R *is a naring, then* $N(R - N(R, F), F) = 0$.

Theorem 3. *If a naring* R *is a subdirect sum of narings* $R_\alpha (\alpha \varepsilon \Omega)$ *and* $N(R_\alpha, F) = 0$ *for every* $\alpha \varepsilon \Omega$, *then* $N(R, F) = 0$.

Theorem 4. *If* R *is a naring, then* $N(R, F) = 0$ *if and only if* R *is isomorphic to a subdirect sum of subdirectly irreducible narings* $R_\alpha (\alpha \varepsilon \Omega)$ *for which* $N(R_\alpha, F) = 0$.

Theorem 5. *A subdirectly irreducible naring* R *has* $N(R, F) = 0$ *if and only if the minimal ideal* J *of* R *contains a nonzero element* a *such that* $F(a) = 0$.

5. The radical of a non-associative ring. In this section we shall discuss the special mapping $a \rightarrow F_1(a)$, where $F_1(a) = I([ax - x + ya - y; x, y \varepsilon R])$. Using the results of Section 2, it is easy to check that this mapping satisfies our basic assumption of Section 4. We shall write N for $N(R, F_1)$ and we shall call N the radical of the naring R. If $N = 0$, we shall say that the naring R is *semi-simple* and if $N = R$ we shall call the naring R a *radical naring.*

Theorem 6. *A subdirectly irreducible naring is semi-simple if and only if it is a simple naring with unit element.*

Proof. Excluding the trivial case of a one-element ring, the direct statement follows from Theorem 5 since $F_1(1) = 0$. Conversely, by Theorem 5, there is a nonzero element e of the minimal ideal J of R for which $F_1(e) = 0$. Then e is the unit element of R and $J = R$ so that R is simple.

Corollary. *A simple naring is semi-simple if it has a unit element, otherwise it is a radical naring.*

Remark. In the associative case, one is able to show that a simple ring with left (right) unit has a unit, and consequently that the G-, F_1-, and

$F_1{}^r$-radicals all coincide. This result is no longer true in the non-associative case, as the following example shows. Let A be the non-associative algebra of order two over a field F with basal units e and u and multiplication defined by $ee = e$, $eu = u$, $ue = 0$, and $uu = e$. This algebra is simple since a nonzero ideal M of A contains a nonzero element $\alpha e + \beta u$ and hence also contains $u(\alpha e + \beta u) = \beta e$. Thus M contains e unless $\beta = 0$, but then α is nonzero and M contains αe, so that $e \varepsilon M$ in any case. Thus $M = A$ and A is a simple non-associative algebra with left unit but no unit. The existence of a unit for simple narings with one sided units may be obtained without the full force of the associative law. We may, for example, assume that association is *symmetric* in the sense that if $a(bc) = (ab)c$, then a, b, c associate in any order. Then if R is a simple naring with a left unit and if association is symmetric in R, R has a unit element. An example of a naring in which association is symmetric is any *alternative* ring.

THEOREM 7. *The radical of a naring R is the intersection of all the ideals M of R such that $R - M$ is a simple naring with unit element.*

COROLLARY 1. *If R is not a radical naring then the radical of R is the intersection of all the maximal ideals M of R such that $R - M$ has a unit element.*

COROLLARY 2. *If R is a naring which is not a one-element ring and has a unit element, then the radical of R is the intersection of all the maximal ideals of R.*

THEOREM 8. *The naring R is semi-simple if and only if it is isomorphic to a subdirect sum of simple narings with unit element.*

THEOREM 9. *If the descending chain condition holds for the ideals of a semi-simple naring R, then R is isomorphic to the full direct sum of a finite number of simple narings with unit elements.*

Remark. We postpone a discussion of Theorem 10 of Brown and McCoy with the remark that the Jacobson radical has been defined only for associative and alternative rings.[7]

THEOREM 11. *If R is a power-associative*[8] *naring, that is, if each element of R generates an associative subnaring of R, and if every element of $I(b)$ is nilpotent, then b is in the radical of R.*

---

[7] N. Jacobson [1] and Smiley [1].

[8] A. A. Albert [2].

THEOREM 12. *If A is an ideal of the naring R, the radical of A is contained in the radical of R.*

COROLLARY. *Any ideal of a semi-simple naring is semi-simple.*

6. Relation to the radical of A. A. Albert. In this and the following section we shall discuss the relation of the radical of a non-associative ring which was defined in Section **6** to previous definitions of the radical of certain non-associative systems given by A. A. Albert[3] and by Max Zorn.[4]

A. A. Albert calls a non-associative algebra A of finite order over a field "semi-simple" in case A is the direct sum of (finitely many) nonzero simple non-associative algebras. Then if A is homomorphic to a "semi-simple" non-associative algebra, Albert definies the radical N' of A to be the intersection of the family of ideals B_α of A for which $A - B_\alpha$ is "semi-simple." We shall show in this section that $N' = N$ when the basic ring R is a non-associative algebra of finite order *with a unit element.* We emphasize that all of our theorems are valid for non-associative algebras.

LEMMA. *If B is an ideal of a non-associative algebra A such that $A - B$ is isomorphic to the direct sum of nonzero simple non-associative algebras C_i $(i = 1, \cdots, n)$, then there are ideals B_i of A $(i = 1, \cdots, n)$ such that each $A - B_i$ is a nonzero simple non-associative algebra and $\Pi B_i = B$.*

Proof. The mapping $a \rightarrow aH_i = c_i$ is a homomorphism of $A - B$ onto C_i with kernel M_i. Now M_i is an ideal of $A - B$ and $(A - B) - M_i \cong C_i$. Each M_i gives rise to an ideal B_i of A such that $B_i \geqq B$ and $M_i = B_i - B$. Hence $A - B_i \cong (A - B) - (B_i - B) = (A - B) - M_i \cong C_i$, so that $A - B_i$ is a nonzero simple non-associative algebra. To see that $\Pi B_i = B$, note that $\Pi M_i = 0$ from which $\Pi B_i = B$ follows readily.

COROLLARY 1. *If A is a non-associative algebra which is homomorphic to a "semi-simple" non-associative algebra, then N' is the intersection of all the ideals C_β of A for which $A - C_\beta$ is a nonzero simple non-associative algebra.*

COROLLARY 2. *If A is a non-associative algebra with unit element then N' is the intersection of all ideals C_α of A for which $A - C_\alpha$ is a simple non-associative algebra with unit element.*

Proof. It suffices to remark that if $A - C_\alpha$ is a one element algebra, then $A = C_\alpha$.

THEOREM 13. *If A is a non-associative algebra of finite order over a field and which has a unit element, then $N' = N$.*

Proof. This is clear by Corollary 2 of the Lemma and Theorem 7.

Albert has given an example[8] for which N' is a *field* provided that the characteristic of the base field is *not two.* In this exceptional case the associative subalgebra spanned by e and u contains the nilpotent ideal generated by $e + u$ and *is not* semi-simple. We find that N' is the ideal of A spanned by $e + u$ and v.

7. The radical of a hypercomplex alternative ring.

We continue our discussion of previous work on the radical of certain non-associative systems. Max Zorn[9] proved that if an alternative ring R was a *hypercomplex* alternative ring, then the set of properly nilpotent elements of R is an ideal and further if this ideal is zero for R, then R is the direct sum of a finite number of simple alternative rings (with unit). We have shown elsewhere[10] that Jacobson's definition of the radical of a ring[8] applies to alternative rings and is the set of all properly nilpotent elements provided that the alternative ring is a hypercomplex alternative ring. In this section we shall show that the radical of Brown and McCoy also reduces to the set of all nilpotent elements for a hypercomplex alternative ring.

A *hypercomplex* alternative ring is an alternative ring which satisfies the following chain conditions.

(CI) *Every sequence $(a^nR; n = 1, 2, \cdots)$ is ultimately constant.*

(CII) *Every monotone sequence $(A_r(a_n); n = 1, 2, \cdots)$ is ultimately constant.*

Here $A_r(a_n) \equiv [x; x \varepsilon R, a_nx = 0]$ is the set of all right annihilators of a_n. We then have the following theorem.

THEORM 10. *If R is a hypercomplex alternative ring, then N is the set of all properly nilpotent elements of R, that is, N is the radical of R in the sense of Zorn.*

Proof. We have shown elsewhere[10] that under the hypothesis of our theorem, the set of all properly nilpotent elements of R coincides with the radical of R as defined by Jacobson. This latter radical consists of all elements $b \varepsilon R$ such that $a \varepsilon I(b)$ implies that there is an element $c \varepsilon R$ so that $a - c + ac = 0$. We have also shown that $a - c + ac = 0$ for a, $c \varepsilon R$ if and only if every element of R is in the set $[ax - x; x \varepsilon R]$. Then clearly N contains the radical N_2 of Zorn.

---

[9] Max Zorn [1] and [2].

[10] Smiley [1].

To prove that the radical of Zorn contains N we use the result of Zorn which states that N_2 is an ideal of R. Observe that $R - N_2$ is a hypercomplex alternative ring and that $N_2(R - N_2) = 0$. Then $R - N_2$ is a direct sum of simple alternative rings with unit elements. Thus Theorem 8 yields $N(R - N_2) = 0$, from which $N_2 \geqq N$ follows easily. The proof is complete.

NORTHWESTERN UNIVERSITY.

BIBLIOGRAPHY.

A. A. Albert:

[1] "The radical of a non-associative algebra," *Bulletin of the American Mathematical Society*, vol. 48 (1942), pp. 891-897.

[2] "Power associative rings. I," (Abstract), *Bulletin of the American Mathematical Society*, vol. 53 (1947), p. 905.

Garrett Birkhoff:

[1] *Lattice Theory*, New York City, 1940.

B. Brown and N. H. McCoy:

[1] "Radicals and Subdirect Sums," *American Journal of Mathematics*, vol. 69 (1947), pp. 46-58.

N. Jacobson:

[1] "The radical and semi-simplicity for arbitrary rings," *American Journal of Mathematics*, vol. 67 (1945), pp. 300-320.

N. H. McCoy:

[1] "Subdirect sums of rings," *Bulletin of the American Mathematical Society*, vol. 53 (1947), pp. 856-877.

M. F. Smiley:

[1] "The radical of an alternative ring," *Annals of Mathematics*, vol. 49 (1948), pp. 702-709.

M. Zorn:

[1] "Theorie der alternativen Ringe," *Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität*, vol. 8 (1930), pp. 123-147.

[2] "Alternative rings and related questions I: Existence of the radical," *Annals of Mathematics* (2), vol. 42 (1941), pp. 676-686.

ON n-ALITY THEORIES IN RINGS AND THEIR LOGICAL ALGEBRAS, INCLUDING TRI-ALITY PRINCIPLE IN THREE VALUED LOGICS.*[1]

By ALFRED L. FOSTER.

Introduction. In every ring $(R, +, \times)$ there exists an intrinsic but usually dormant duality-symmetry theory which specializes to the familiar Boolean duality when R is a Boolean ring. This theory has been presented and developed in diverse directions in a series of papers [1], $\cdots$, [7],[2] with several of which it will be necessary to establish contact in the present communication.

It was later discovered, as first broadly outlined in a portion of [2], that this duality theory of rings is itself but an instance of a host of K-ality theories, based on certain preassigned groups K, and that these in turn constitute merely one class of realizations of a general transformation theory,—a simple unifying skeletal framework which also includes traditional transformation and invariant theories among its specializations.

Even on the *simple* (later also called mod C) duality level, as we now refer to the original ring duality to distinguish it from rival theories, we were able to formulate numerous interesting concepts, such as the (simple) 'logical algebra' of a ring, and to profitably explore such questions as the strength of the bond between a ring and its (simple) logic; these include generalizations of familiar Boolean questions. (See especially [1]). In the present communication we (a) elaborate the general transformation theory and employ it to (b) elevate various 'logical algebra' concepts from the simple to a much more general level. In so doing (c) the special role of the simple level is considerably illuminated.

In **6** we (d) put forward the concept 'p-ring' as a natural generalization of Boolean rings (which latter are identical with 2-rings),[3] and particularly

---

* Received November 1, 1948.

[1] A segment of this paper was presented to the National Academy of Sciences, Berkeley, Nov. 1948.

[2] Numbers in square brackets refer to the bibliography given at the conclusion.

[3] The concept "p-ring" was first defined by McCoy and Montgomery, in [9]—for which reference I am indebted to the referee. Strictly, our "p-ring" is the "p-ring with unit" of [9].

study the case $p = 3$. It is shown that each 3-ring is interdefinably bound to its 'logical algebra' in a manner which generalizes the familiar interdefinable bond between a Boolean ring (i. e., $p = 2$) and its Boolean (= logical) alegbra. Each 3-ring-algebra is shown to possess an intrinsic *tri-ality* theory, the successor of the familiar Boolean duality case for $p = 2$. In this way the same tight intimacy which exists between the logic of propositions (= 2-valued logic), Boolean rings, Boolean algebras and the omnipresent Boolean duality theory on the one hand, is shown to extend to the 3-valued logic, 3-rings, the corresponding 3-algebras, and the engulfing tri-ality theory on the other.

For arbitrary p (= prime), we exhibit a *p-ality* theory connecting each *p-ring* with its logical alegbra. For $p > 3$, however, no formula has as yet been found which equationally defines a p-ring in terms of its logical algebra, in fact the existence of such a formula has not been settled,—(see **8-10**).

1. (Simple) Duality theory of rings.

To facilitate reference and orientation, in this section we very briefly recall a few essentials of the simple duality theory of rings.

Let $R = (R, +, \times)$ be a ring with a unit element. Each concept in R is shown to possess a dual concept; in particular: 0 and 1 are dual elements; $\times, \otimes$; $+, \oplus$; $-, \ominus$; * are dual operations, the latter being self-dual, where

(1. 1) $a \otimes b = a + b - a \times b, \quad a \times b = a \oplus b \ominus a \otimes b$ = dual ring products

(1. 2) $a \oplus b = a + b - 1, \quad a + b = a \oplus b \ominus 0$ = dual ring sums

(1. 3) $a \ominus b = a - b + 1, \quad a - b = a \ominus b \oplus 0$ = dual ring differences

(1. 4) $a^* = 1 - a = 0 \ominus a =$ (self-dual) ring complement.

Restricted for brevity to these operations the duality theorem reduces to:

(SIMPLE) DUALITY THEOREM FOR RINGS. *If* $P(0, 1; \times, \otimes; +, \oplus; -, \ominus; ^*)$ *is a true proposition of a given ring* R, *so also is its* (*simple*) *dual*, $\mathrm{dl}\, P = P(1, 0; \otimes, \times; \oplus, +; \ominus, -; ^*)$ *obtained by replacing each argument by its dual, with* * *left unchanged* (*self dual*).

The duality theorem is illustrated by each of the relations (1. 1)-(1. 4). Also by the dual theorems, holding in any ring,

$$(1.5)\qquad (a\times b)^* = a^*\otimes b^*, \qquad (a\otimes b)^* = a^*\times b^*$$
$$= \text{Ring 'De-Morgan' formulas,}$$

and the self dual

$$(1.6)\qquad a^{**} = a, \qquad 0^* = 1, \qquad 1^* = 0.$$

Again, by

$$(1.7)\quad \begin{array}{ll} (R, +, \times) & \text{is a ring with 0 as zero element and 1 as unit.} \\ (R, \oplus, \otimes) & \text{is a ring with 1 as zero element and 0 as unit.} \end{array}$$

Also, in any Boolean-like ring, (see [1]),

$$(1.8)\quad a + b = (a\times b^*)\otimes(a^*\times b), \qquad a\oplus b = (a\otimes b^*)\times(a^*\otimes b).$$

Again, in any field $(F, +, \times)$, (see [6]),

$$(1.9)\quad \begin{array}{ll} a\times a^{01}\times a^{10} = \mu = \text{constant} \neq 0 & (a\neq 0, \neq 1) \\ a\otimes a^{10}\times a^{01} = \tau = \text{constant} \neq 1 & (a\neq 1, \neq 0) \end{array}$$

$$(1.10)\quad a + b : \begin{cases} a + b = a\otimes b\otimes (a^0\times b^0) & (a\neq 1, b\neq 1) \\ a + 1 = 1 + a = a\otimes a^{10} & (a\neq 0, \neq 1) \\ 1 + 1 = \tau, \qquad 1 + 0 = 0 + 1 = 1 & \end{cases}$$

$$(1.11)\quad a\oplus b : \begin{cases} a\oplus b = a\times b\times (a^1\otimes b^1) & (a\neq 0, b\neq 0) \\ a\oplus 0 = 0\oplus a = a\times a^{01} & (a\neq 1, \neq 0). \\ 0\oplus 0 = \mu, \qquad 0\oplus 1 = 1\oplus 0 = 0 & \end{cases}$$

In (1.9) and (1.10), a^1 *is the* $\times$ inverse of a, and a^0 the $\otimes$ inverse. These few illustrations will suffice for our purpose. In case R is a Boolean ring we recall that: the dual ring products $\times$, $\otimes$ respectively reduce to the usual Boolean logical product, $\cap$, and logical sum, $\cup$, the ring complement $*$ to the Boolean complement, $\neg$, and the ring 'De Morgan' and duality theorems to the corresponding familiar Boolean theorems.

With the above Boolean specialization as motivation, in any ring R the (operationally closed) system $(R, \times, \otimes, {}^*)$ was introduced in [1] as the (simple) *logical algebra* (also briefly as the *Logic*) of the ring. We recall that a *Logically definable* ring is one whose ring $+$, and consequently the entire ring, is definable in terms of its Logic, as is the case for instance in a Boolean ring. These notions will be reexamined and refined in **5**.

In addition to the usual $(R, +, \times)$ notation for a ring we may also write it in the (simple) dual form $(R, \oplus, \otimes)$, or in the 'mixed' form $(R, +, \oplus, \times, \otimes, {}^*)$, etc. Later, corresponding to other duality or, more generally, n-ality theories, we have n pure forms

$$(R, +, \times), \qquad (R, +', \times'), \qquad (R, +'', \times''), \cdots$$

and corresponding 'mixed' forms

$$(R, +, +', +'', \cdots, \times, \times', \times'', \cdots, {}^{0}, {}^{0'}, {}^{0''}, \cdots).$$

2. Perspective of the general transformation theory. In the traditional applications of the usual transformation and invariant theories to various mathematical disciplines, basically one is concerned with a set (generally a group) of admissible 'coordinate transformations,' and with the changes suffered by (or with the invariance of) certain mathematical concepts of the discipline when one passes from one admissible coordinate system to another such. In all of the classical applications the underlying 'computational disciplines' (arithmetic, various algebras, analysis etc.) are absolute invariants (scalars), i. e., unchanged by any of the coordinate transformations. In [2] it was first sketched how this transformation-invariant theory may be profitably extended to permit much wider applications, as for instance to the above types of computational disciplines themselves. Thus one is led to the conception of arithmetic, or analysis, or a particular ring or class of rings, in fact any kind of operational algebra, as a discipline whose concepts transform 'cogrediently,'—or sometimes 'contragrediently' with each permissible change of 'coordinates,' in a manner analogous to (and possessing as a particular specialization) the transformation of tensors or matrices to new coordinates.

In general one is forced to deal with 'mixed' as well as 'pure' notions, the latter being such as are defined entirely within a single coordinate system, while the former are defined in terms of at least two permissible systems. Thus, for example, while a ring defined in either of the (simple) dual forms $(R, +, \times)$ or $(R, \oplus, \otimes)$ is a pure concept, the Logic $(R, \times, \otimes, {}^{*})$ of the ring is perforce a 'mixed' notion.

For each given group K of permissible 'coordinate transformations' in a discipline there is a *K-ality theorem* which explicitly formulates the manner in which the true propositions of the discipline transform when passing from one permissible coordinate system to another. Thus for instance, corresponding to each group K of order 2 one has an accompanying duality theory; in particular, applied to a ring R and with K chosen as the *complementation group* $C = C(R)$, of order 2,

$$(2.1) \qquad x^{*} = 1 - x, \qquad x^{**} = x = \text{identity},$$

the corresponding ring duality is the *simple* theory partially recalled in **1.**

From the general transformation theory point of view, then, the 'K-ality' of concepts,—also variously referred to as 'n-ality, mod K,' or 'n-ality (K),' etc., if K is a finite group of order n, is simply a way of saying that the concepts are identical, but expressed in different coordinates. Thus for example, in the (simple) ring duality theory specialized to Boolean rings, the familiar logical product $\times\ (=\bigcap)$ and logical sum $\otimes\ (=\bigcup)$ are exhibited as the *same* concept, expressed in different coordinates.

With the exception of [4] and parts of [2], the narrow band of applications of this transformation theory explored in the series $[1], \cdots, [7]$ is more or less identified with rings, and with the permissible group K taken as the simple complementation group C. We shall here still be concerned with rings, but not alone with this special choice of the group K.

3. General transformation theory (continued). Let $U = \{\cdots, x, \cdots\}$ be a class (with or without structure), and $\Phi_U = \Phi = \{\cdot \cdot \cdot, \phi, \cdot \cdot \cdot\}$ the set of all operations (or *multitations,*—see [4])

$$(3.1) \qquad \phi = \phi(x, \cdot \cdot \cdot) = \varepsilon\, U$$

(or one or more arguments) of U into itself. A ϕ of a single argument is also called a *monotation,* $\phi(x)$; similarly for *bitations* $\phi(x, y)$, etc. We have

$$(3.2) \qquad \Phi = \sum_{i=1,2,\ldots} \Phi^{(i)}$$

where $\Phi^{(i)}$ is the class of all i-tations of U, and $\sum$ denotes set union. A permutation ρ is a 1-1 reversible monotation,[4] whose inverse we write $\bar{\rho}$. For $\sigma \varepsilon \Phi^{(i)}$ and $\phi \varepsilon \Phi$ we denote composition (= composite product) by simple juxtaposition,

$$(3.3) \qquad \sigma\phi = (\sigma\phi)(x, \cdot \cdot \cdot) = \sigma(\phi(x, \cdot \cdot \cdot)).$$

We recall the well known associativity of composition for monotations,

$$(3.4) \qquad \sigma(\sigma'\sigma'') = (\sigma\sigma')\sigma'' = \sigma\sigma'\sigma''.$$

Solely in the interest of convenient distinction we sometimes refer to the elements of the set U as 'points,' and those of the set Φ_U as Points (capital P): $U =$ 'point space,' $\Phi =$ 'Point space.'

---

[4] In particular, U need not, of course, be infinite.

Each point permutation ρ induces a Point monotation, defined by

$$\phi \to \phi_\rho = \rho^{-}\phi(\rho(x), \rho(y), \cdots). \tag{3.5}$$

The Point ϕ_ρ we call the *transform of* ϕ *by* ρ.

Thus, if U is chosen as a linear vector space $U = \{\cdots, \bar{x}, \cdots\}$, ϕ as a linear map

$$\phi(\bar{x}) = \begin{pmatrix} \phi_{11} & \phi_{12} \cdots \\ \phi_{21} & \phi_{22} \cdots \\ \vdots & \vdots \end{pmatrix}$$

and ρ as a non-singular linear map, the transform (3.5) reduces to the familiar matrix transform $\rho^{-}\phi\rho$. Similar familiar interpretations result if U is taken as an abstract group. Again if U is taken as a ring $(U, +, \times)$ with unit, ϕ as the ring product $\times$,

$$\phi(x, y) = x \times y \tag{3.6}$$

and ρ as the *complementation* permutation

$$\rho(x) = x^* = 1 - x = \rho^{-}(x), \tag{3.7}$$

the ρ transform of $\times$ is the simple dual, $\otimes$, given by (1.1); and quite generally, the ρ transform of any concept of U is its simple dual. (Compare with **1**).

THEOREM 1. (a) *Each induced Point monotation* (3.5) *is a Point permutation.* (b) *If* $K = \{\cdots, \rho, \cdots\}$ *is a group of point permutations, the corresponding set of induced Point permutations* (3.5) *form a group, which is isomorphic with the group* K. *One has*

$$\{\phi\}_{\rho\rho'} = \{\phi_\rho\}_{\rho'}. \tag{3.8}$$

Proof of (a). Here we must show that (a′): (3.5) is a univoque Point monotation, i. e.,

$$\phi \neq \psi \to \phi_\rho \neq \psi_\rho, \tag{a′}$$

and secondly (a″): for each Point ψ there exists a Point ψ' such that $\psi'_\rho = \psi$. If (a′) were false there would exist points $x_0, y_0, \cdots$ such that

$$\phi(x_0, y_0, \cdots) \neq \psi(x_0, y_0, \cdots) \tag{3.9}$$

$$\rho^{-}\phi(\rho(x), \rho(y), \cdots) \underset{x, y, \cdots}{\equiv} \rho^{-}\psi(\rho(x), \rho(y), \cdots). \tag{3.10}$$

This is impossible since (3.10) implies

$$\phi(\rho(x),\rho(y),\cdots) \underset{x,y,\cdots}{\equiv} \psi(\rho(x),\rho(y),\cdots), \tag{3.11}$$

which is in contradiction with (3.9), as is seen by taking $x=x_1$, $y=y_1,\cdots$ where

$$\rho(x_1)=x_0, \qquad \rho(y_1)=y_0,\cdots. \tag{3.12}$$

This proves (a′). Again, for given $\psi\,\varepsilon\,\Phi$, (a″) may be satisfied by defining ψ' by

$$\psi'(x,y,\cdots)=\rho(\psi(\rho^-(x),\rho^-(y),\cdots)). \tag{3.13}$$

This completes part (a).

The relation (3.8) follows directly from the usual expression for the inverse of a product:

$$\begin{aligned}\{\phi_\rho\}_{\rho'}(x,y,\cdots)&=\rho'^-(\rho^-(\phi(\rho(\rho'(x)),\rho(\rho'(y)),\cdots)))\\ &=(\rho\rho')^-(\phi(\rho\rho'(x),\rho\rho'(y),\cdots))=\phi_{\rho\rho'}.\end{aligned} \tag{3.14}$$

Part (a) and (3.8) together show that the set of induced Point monotations form a group which is a homomorphic image of the group K. That we actually have an isomorphic image may be seen as follows. One must show that for $\rho\neq\rho'$, (b′): $\phi\rightarrow\phi_\rho$ and $\phi\rightarrow\phi_{\rho'}$ are different Point permutations. From the premise, for some $x_0\,\varepsilon\,U$,

$$\rho(x_0)\neq\rho'(x_0). \tag{3.15}$$

Denote

$$\rho(x_0)=x_1, \qquad \rho'(x_0)=x_2. \tag{3.16}$$

Then $x_1\neq x_2$. Consider the Point $\psi(x)\underset{x}{\equiv}x_1$. One has

$$\begin{aligned}\psi&\rightarrow\psi_\rho=\rho^-\psi(\rho(x))\underset{x}{\equiv}\rho^-(x_1)=x_0\\ \psi&\rightarrow\psi_{\rho'}=\rho'^-\psi(p'(x))\underset{x}{\equiv}\rho'^-(x_1).\end{aligned} \tag{3.17}$$

Now $\rho'^-(x_1)\neq x_0$, since

$$\rho'^-(x_1)=x_0\rightarrow x_1=\rho'(x_0)=x_2, \tag{3.18}$$

contrary to hypothesis. Hence (b′) is proved, and with it the Theorem.

The group of Point transformations induced by a group K will not in general be transitive, even if the point transformation group is transitive.

The totality of Points ϕ_ρ into which a given Point ϕ is transformed by the various $\rho \varepsilon K$ is called a *congruence class* (of Points), mod K. We write

$$\phi \equiv \phi'(K) \tag{3.19}$$

to denote that ϕ and ϕ' are in the same congruence class. Since each congruence class mod K forms a Point set on which the group K is homomorphically represented as a transitive permutation group, by a well known theorem on the degree of a transitive permutation group one has the

THEOREM 2. *If the group K is finite, of order n, the number of Points in a congruence class, mod K, is a divisor of n.*

In particular, if K is a (cyclic) group of prime order p, the number of Points in each congruence class mod K is either p or 1.

Since $K \subset \Phi$, each $\rho \varepsilon K$ is of course a Point. One has

THEOREM 3. *If K is an Abelian group, each $\rho \varepsilon K$ is fixed under K, i. e., forms a congruence class of a single Point.*

The proof is immediate from (3.5). The self-duality of the ring complement, $*$, in the simple ring duality theory, and similarly the self-tri-ality of the cyclical negation, $^\wedge$, as well as its inverse, $^\vee$, in the tri-ality theory of 3-rings considered in **6**-**10**, is an immediate consequence of Theorem 3. Of course there may also be fixed Points ϕ, mod K, which are not elements of K.

It is often advantageous to regard these general transformation notions in a different light, corresponding to **2**. We may look on K as a group of permissible 'coordinate transformations' in U. In the 'ρ coordinate system' the point x receives the new 'coordinate' $\rho(x)$ [5]; the multiplication ϕ 'becomes' ϕ_ρ. That is, ϕ_ρ, which is an isomorphic image of ϕ by Theorem 1, is the 'same' multitation as ϕ, described however in the ρ coordinate system.

For a given group K of permissible coordinate transformations we also say: ϕ and ϕ' are 'K-als,' instead of $\phi \equiv \phi'(K)$. Moreover if K is finite, of order n, we also speak of 'n-als' mod K; for $n = 2, 3, \cdots$ we speak of 'duals,' 'tri-als,' etc. A fixed Point ϕ is then called 'self-K-al,' respectively 'self-dual,' 'self-tri-al,' etc.

If $A = \{\alpha_1, \alpha_2, \cdots\}$ is any subset of Φ_U, we denote by $|A|$ the subset of Φ_U which is compositionally generated by the $\alpha_i \varepsilon A$. For instance if

---

[5] More accurately, by the 'point x' we mean the point which, in the ξ coordinate system, has the coordinate x, where ξ is the identity of the group K.

$A = \{\alpha, \beta\}$, and if α, β are of the form $\alpha(x), \beta(x, y)$, then $|A|$ would contain the multitations

$$(3.20) \qquad \alpha(\alpha(x)), \alpha(\beta(x,y)), \beta(\alpha(x), \beta(y,z)), \beta(\alpha(x), \beta(x,y)), \text{ etc.}$$

We have of course

$$(3.21) \qquad A \subseteq |A| \subseteq \Phi\Phi_U.$$

4. **K-logical definability of rings.** We here clarify the concept of the simple logical (algebra) definability of rings (see end of **1**), and lift it from the simple to the general level. We specialize: $U = R = (R, +, \times)$ is a ring (which need not contain a unit). Let

$$(4.1) \qquad K = \{\xi, \rho', \rho'', \cdots\} = \{\cdots, \rho, \cdots\}$$

be a group of coordinate transformations in (i. e., permutations of) the class R, with the identity of K denoted by ξ, and let

$$(4.2) \qquad \{\times, \times', \times'', \cdots\} = \{\cdots, \times_\rho, \cdots\}$$

be the class of all transforms of the ring product $\times$ by the various $\rho \varepsilon K$. (This congruence class mod K is evidently the same as the class of transforms of any fixed $\times^{(i)}$). Here, by (3.5),

$$(4.3) \qquad a \times_\rho b = \rho^-(\rho(a) \times \rho(b)),$$

and obviously

$$(4.4) \qquad a \times_\xi b = a \times b.$$

The algebra

$$(4.5) \qquad (R, \times, \times', \times'', \cdots, \xi, \rho', \rho'', \cdots)$$

whose class, R, is taken as identical with that of the given ring $(R, +, \times)$, and whose operations are $\times, \times', \times'', \cdots, \xi, \rho', \rho'', \cdots$ as indicated, we call the *logical alegbra* (mod K of the ring, or simply the *K-logic* of the ring. In addition to (4.5) we also denote this K-logic simply by

$$(4.6) \qquad (R, \times, K) = (R, \times', K) = (R, \times'', K), \text{ etc.}$$

R is closed with respect to each of the operations $\times, \times', \times'', \cdots \xi, \rho', \rho'', \cdots$. A K-logical concept of a ring $(R, +, \times)$ is one which is definable entirely in terms of the K-logic of the ring. Among the K-logical concepts of a ring are of course all multitations belonging to the class $|\times, K|$, i. e., all multitations compositionally generable from $\times$ and the $\rho \varepsilon K$. We note that

$$(4.7) \qquad |\times, K| = |\times', K| = \cdots = |\times, \times', \times'', \cdots, \xi, \rho', \rho'', \cdots|.$$

Further, if $\rho_1, \rho_2, \cdots$ are a set of generators of the group K one has

$$(4.8) \qquad |\times, K| = |\times, \rho_1, \rho_2, \cdots|.$$

In [1] and also in [6] special cases of the K-logical definability of the ring sum, $+$, — and hence of the whole ring, were treated. We here generalize and simultaneously clarify these applications by introducing several refinements. We say that a ring $(R, +, \times)$ is: (a) K-logically *definable* if its $+$ is a K-logical concept of the ring. (b) K-logically *equationally* definable if its $+$ is $\varepsilon\,|\times, K|$, i. e., if its $+$ satisfies an identity

$$(4.9) \qquad a + b \underset{a,b}{\equiv} \phi(a, b) \text{ where } \phi \,\varepsilon\, |\times, K|.$$

(c) K-logically *fixed*, if it is K-logically definable, and if there exists no other ring $(R, +_1, \times)$,—on the same set R and with the same product $\times$, but with $+_1 \neq +$, which is K-logically definable.

We shall illuminate these distinctions, and at the same time establish their essential independence, by again returning to the *simple* or C-logics (C = complementation group (2.1)).

A Boolean ring is C-logically equationally definable, since in such a ring (see [1] and 1),

$$(4.10) \qquad a + b = ab^* \otimes a^*b.$$

It is easy to extend this to the

THEOREM 4. *A Boolean ring is C-logically equationally definable and fixed.*

Proof. Let $(R, +, \times)$ be a Boolean ring, and let $(R, +_1, \times)$ be a ring having the same C-logic as $(R, +, \times)$. By Stone's theorem [8], a Boolean ring is characterized by the idempotency condition

$$(4.11) \qquad x \times x = x^2 = x \qquad (x \,\varepsilon\, R)$$

from which follows that

$$(4.12) \qquad x + x = 0.$$

Hence $(R, +_1, \times)$ is also Boolean, and

$$x +_1 x = 0. \tag{4.13}$$

In addition, by hypothesis,

$$1 - x = 1 -_1 x = x^*. \tag{4.14}$$

Now from (4.11) and (4.12), respectively (4.13), follows

$$x + y = x +_1 y \underset{x,y}{\equiv} xy^* \otimes x^*y, \tag{4.15}$$

which proves the theorem.

We know also that the more general class of *Boolean-like* rings (see [1]) are C-logically equationally definable, with $a + b$ again given by (4.10); in fact this was essentially the definition of this class of rings. In contrast to Theorem 4, however, we have

THEOREM 5. *A Boolean-like ring is C-logically equationally definable, but not in general C-logically fixed.*

Proof. Consider the two rings $(R, +, \times)$ and $(R, +_1, \times)$, where $R = \{0, 1, 2, 3\}$ and where

$$\begin{array}{c|cccc} \times & 0 & 1 & 2 & 3 \\ \hline 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 2 & 3 \\ 2 & 0 & 2 & 0 & 2 \\ 3 & 0 & 3 & 2 & 1 \end{array} \qquad \begin{array}{c|cccc} + & 0 & 1 & 2 & 3 \\ \hline 0 & 0 & 1 & 2 & 3 \\ 1 & 1 & 0 & 3 & 2 \\ 2 & 2 & 3 & 0 & 1 \\ 3 & 3 & 2 & 1 & 0 \end{array} \qquad \begin{array}{c|cccc} +_1 & 0 & 1 & 2 & 3 \\ \hline 0 & 0 & 1 & 2 & 3 \\ 1 & 1 & 2 & 3 & 0 \\ 2 & 2 & 3 & 0 & 1 \\ 3 & 3 & 0 & 1 & 2 \end{array}. \tag{4.16}$$

Here $(R, +, \times) = H_4$, the simplest Boolean-like ring which is not also Boolean (see [1]), while $(R, +_1, \times) = ((4))$ is the ring of residues mod 4. From (4.16) it is easily found that whether $*$ be computed from H_4 or from $((4))$, the results are identical,

$$x^* = 1 - x = 1 -_1 x; \qquad * \frac{0\ 1\ 2\ 3}{1\ 0\ 3\ 2}. \tag{4.17}$$

Hence we have two distinct (even non-isomorphic) rings having identical C-logics, which proves Theorem 5.[6]

It was shown in [6] that a field $(F, +, \times)$ is always C-logically definable,—one such definition is recalled in (1.10) of **1**. While this particular C-logical definition of $+$ is obviously not equational, an equational one might conceivably exist. We show

---

[6] We have, of course, simultaneously shown that $((4))$ is not C-logically fixed.

Theorem 6. (a) *A Field* $(F, +, \times)$ *is a C-logically fixed ring. However* (b) *in general a field will have no C-logical equational definition.*

Proof of (a): We first prove the

Lemma. *Let* $(R, +, \times)$ *and* $(R, +_1, \times)$ *be rings, and let* $(R, +, \times)$ *possess no 0-divisors. Then* $(R, +_1, \times)$ *has no 0-divisors, and*

$$-x = -_1 x \qquad (x \varepsilon R). \tag{4.18}$$

Proof. Since $x(-1) = -x$, $x(-_1 1) = -_1 x$, it is obviously sufficient to show that (A): $-1 = -_1 1$. We have

$$\{(-1)(-_1 1)\}\{(-1)(-_1 1)\} = (-1)^2(-_1 1)^2 = 1, \tag{4.19}$$

since -1 commutes with all elements. Since no 0-divisors exist, $x^2 = 1 \rightarrow x = 1$ or $x =$ additive inverse of 1. Hence

$$(-1)(-_1 1) = 1 \text{ or } -1 \text{ or } -_1 1. \tag{4.20}$$

If $(-1)(-_1 1) = 1$, then $(-_1 1) = -1$, and (A) is proved. If $(-1)(-_1 1) \neq 1$, then since $\times$ is the same in both rings,

$$(-1)(-_1 1) = -1 = -_1 1. \tag{4.21}$$

Hence (A) and with it the Lemma is proved.

Note. A comparison of H_4 and ((4)) shows that the Lemma is in general false if 0-divisors are present.

The proof of part (a) of Theorem 6 is now immediate. Let $(F, +, \times)$ be a field and let $(F, +_1, \times)$ be a ring having the same C-logic as $(F, +, \times)$. We must show that $+ = +_1$. By hypothesis

$$1 - x = 1 -_1 x. \tag{4.22}$$

By the Lemma we then have

$$x + 1 = x +_1 1. \tag{4.23}$$

We must show that

$$x + y = x +_1 y. \tag{4.24}$$

For $x = 0$ this is trivial, and for $x \neq 0$, by (4.23),

$$x + y = x(1 + yx^{-1}) = x(1 +_1 yx^{-1}) = x +_1 y, \tag{4.25}$$

where x^{-1} is the $\times$ inverse of x. This proves part (a) of Theorem 6.

Proof of part (b). We shall show that $(F_3, +, \times)$, the field of residues mod 3, is not equationally C-logically definable. We recall a special instance of a well known theorem:

(B) Each of the 3^3 monotations $\alpha(x)$ of the class F_3 may be (uniquely) expressed in the 'analytical' form,

$$\alpha(x) = a_0 + a_1x + a_2x^2 \qquad (a_0, a_1, a_2 = 0, 1, 2 \pmod 3)). \tag{4.26}$$

We next show

(C) If $\phi(x)$ is a monotation of F_3 which is $\varepsilon \mid \times, C \mid$, then if $\phi(x)$ is expressed in the canonical form (4.26), $a_0 \neq 2$.

The truth of (C) is seen as follows. The class Γ consisting of all elements of $\mid \times, C \mid$ which are monotations, is inductively defined as follows: 1°) $x \,\varepsilon\, \Gamma$; 2°) if $\sigma(x)$ and $\tau(x)$ are ε's Γ, so are $\sigma^*(x)$, $\sigma^2(x)$, $\sigma(x) \times \tau(x)$. That we need go no higher than σ^2 follows from the identity

$$x^3 \equiv x \qquad (x \,\varepsilon\, F_3). \tag{4.27}$$

We see that x satisfies (C). Also, if $\sigma(x)$ and $\tau(x)$ each satisfy (C), so do σ^2, σ^* and $\sigma \times \tau$, the first and last since the constant term of the product $\sigma \times \tau$ is the product of the constant terms, and the second since $1^* = 0$, $0^* = 1$. Hence, by induction, (C) is proved.

We further observe that in F_3

$$1 \equiv (x \times x^{2*})^* = \varepsilon\, \Gamma. \tag{D}$$

Suppose now that

$$x + y \text{ is } \varepsilon \mid \times, C \mid .$$

Then by (D), $1 + 1 = 2$ would be $\varepsilon \mid \times, C \mid$. This is however impossible by (C), which contradiction completes part (b), and with it Theorem 6.

By considering the prime subfield of a given field one can, with only minor modifications, strengthen part (b) of Theorem 6 to

THEOREM 7. *If F is a field of characteristic $\neq 2$, it cannot be equationally C-logically defined.*

Similarly one may show

THEOREM 8. *The ring $(W, +, \times)$ of whole numbers cannot be equationally C-logically defined.*

5. Ring-logics (K). Let $(R, +, \times)$ be an arbitrary but fixed ring, and K a group of permutations of R. We say that the group K is *semi-adapted* to the ring R if the ring is K-logically fixed; if in addition the ring is equationally definable in terms of its K-logic, we say that K is *fully-adapted* to the ring. We have just seen, for instance, that the simple complementation group C is always fully adapted to Boolean rings, but in general only semi-adapted to a given field.

If K is fully adapted to a ring R, we shall also refer to R as a *ring-logic* (K), or a *logic-ring* (K); in this case ring and K-logic uniquely and equationally fix each other, and it is therefore appropriate to speak of *the* ring of the K-logic, as well as the K-logic of the ring.

It is natural to inquire: Given a ring R, does there always exist at least one group K which is (a) semi-adapted, or (b) fully-adapted to R? Question (a) may be affirmatively answered by a simple construction into which we shall not here enter. Entirely different in nature is the stronger question (b), to which no complete answer has as yet been found. This latter question (b) may be restated: May any ring be converted into a ring-logic (K) by suitably choosing K?

That Boolean rings are not the only ring-logics we shall explicitly show in **8**. In general we may anticipate that the K-ality theory (see **2**) of a ring-logic (K) will be a combinatorially rich theory, in view of the unique equational determinancy of ring in terms of logic and conversely.

6. p-rings. In seeking interesting ring-logics other than Boolean rings we are led to a natural generalization of this latter class.[7] Let p be an arbitrary fixed prime integer. By a *p-ring* we mean a commutative ring with unit $(S, +, \times)$, in which, for all $a \varepsilon S$,

$$a^p = a \times a \times a \times \cdots \times a = a \tag{6.1}$$

$$p^a = a + a + a + \cdots + a = 0. \tag{6.2}$$

The class of Boolean rings is thus coextensive with the class of 2-rings ($p = 2$). For this special case, $p = 2$, (6.2) is a consequence of (6.1), as is well known. That this is not so in general is seen from ((6)), the ring of residues mod 6, in which (6.1) is satisfied with $p = 3$, but not (6.2). It is also easily shown that the prime p is unique for a given p-ring, i. e., that a ring cannot be both a p-ring and a p'-ring, with $p' \neq p$.

---

[7] As already noted, "p-rings" were first introduced in [9], where a proof of Theorem 9 is given.

It is not our purpose here to enter into the structure, either elementary or ideal, of p-rings; we mention however in passing the following extension of a familiar Boolean theorem as

THEOREM 9. *For given prime p, F_p = field of residues mod p is a p-ring, and is a sub-ring of any p-ring. If S is a finite p-ring,*

$$S = F_p \times F_p \times \cdots \times F_p, \tag{6.3}$$

where $\times$ denotes direct product; and hence the number of elements in a finite p-ring is always a power of p, p^t.

Let $S = (S, +, \times)$ be a p-ring. By the *cyclic* (*negation*)[8] group N of S we understand the group of coordinate transformations in S generated by $^\wedge$, where

$$a^\wedge = 1 + a. \tag{6.4}$$

(Here the order p of the cyclic group N is the (prime) characteristic p of N, unlike the complementation group C which always has the fixed order 2).

If $\times, \times', \times'', \cdots$ denote the transforms of $\times$ by $^\wedge$, respectively by $^{\wedge\wedge}$, etc., and if $+', +'', \cdots$ and $-', -'', \cdots$ have similar meanings, by (3.5), (6.4) one easily computes,

$$a \times^{(r)} b = a \times b + r(a + b) + r^2 - r. \tag{6.5}$$

$$a +^{(r)} b = a + b + r. \tag{6.6}$$

$$a -^{(r)} b = a - b - r. \tag{6.7}$$

Again, by Theorem 3, the operation $^\wedge$ is N-fixed, i. e., the same in each permissible coordinate system; this applies also to each of the operations $^{\wedge\wedge}, {}^{\wedge\wedge\wedge}, \cdots$.

In a formula such as (6.4) the coefficient r (mod p) in $r(a + b)$ is an *apparent* (or removable) constant, since it represents $(a + b) + (a + b) + \cdots + (a + b)$. By contrast, the 'additive' constants, such as $r^2 - r$ (mod p) in (6.5), are *real* constants.

We now state, without proof, the p-ality theorem,—the generalization to p-rings of the classic Boolean duality. The proof of the theorem is much like that of the simple duality theory, and offers no difficulty.

THEOREM 10. *p-Ality Theorem for p-Rings. Let S be a p-ring, and let*

---

[8] The terminology 'cyclic negation' is borrowed from the expression by the same name in many-valued logics. See **8, 9**.

$$P(0, 1, 2, \cdots, p-1; \times, \times', \times'', \cdots; +, +', +'', \cdots; \\ -, -', -''; {}^\wedge; {}^{\wedge\wedge}; {}^{\wedge\wedge\wedge}; \cdots)$$

be any true proposition in S involving no apparent constants. Then each of the p-al propositions

$$P' = P(p-1, 0, 1, 2, \cdots; \times', \times'', \cdots; +', +'', \cdots; \\ -', -'', \cdots; {}^\wedge; {}^{\wedge\wedge}; {}^{\wedge\wedge\wedge}; \cdots)$$

$$P'' = P(p-2, p-1, 0, 1, 2, \cdots; \times'', \times''', \cdots; +'', +''', \cdots; \\ -'', -''', \cdots; {}^\wedge; {}^{\wedge\wedge}; {}^{\wedge\wedge\wedge}; \cdots), \text{ etc.}$$

obtained by (a) *leaving each of the operations* ${}^\wedge; {}^{\wedge\wedge}; \cdots$ *unchanged,* (b) *applying any cyclic permutation 'cogrediently' to all other operations and* (c) *the 'contragredient'* (= *inverse*) *permutation to the real constants, is again a true proposition of the ring S.*

The 'contragredient' element of the theorem is one not apparent in the simple case, $p = 2$, since in this case there is no difference between $x^\wedge$ and its inverse, $x^\wedge = x^* = x^\vee$. The self-p-ality of the 'cyclic negation' ${}^\wedge$, as well as that of ${}^{\wedge\wedge}; {}^{\wedge\wedge\wedge}; \cdots$ etc., is a consequence of Theorem 3.

7. 3-ring-logics. We shall now specialize to the case of 3-rings ($p = 3$). Here we explicitly show that the cyclic negation group N (of order 3) is fully adapted to this class of rings, and hence that each 3-ring is a ring-logic (N). The class of 3-ring-logics contains the 3-valued logic ($= F_3$) as its simplest representative, and the 3-valued logic and the general 3-ring-logics are related to each other and dominated by the *tri-ality theory,* exactly as are the ordinary logic of propositions ($= F_2$), 2-ring-logics (= Boolean rings = Boolean algebras) and the enfolding simple (Boolean) duality theory. We may, in this sense, speak of a unified theory. (See **8**, **9**).

Let S be a 3-ring. We have then

$$(7.1) \qquad a^\wedge = 1 + a, \qquad a^{\wedge\wedge} = 1 + 1 + a = 2 + a, \qquad a^{\wedge\wedge\wedge} = a.$$

Let us abbreviate

$$(7.2) \qquad a^{\wedge\wedge} = a^\vee.$$

We have

$$(7.3) \qquad a^\wedge = a^{\vee\vee}; \; a^{\wedge\wedge\wedge} = a^{\vee\vee\vee} = a^{\wedge\vee} = a^{\vee\wedge} = a;$$

$$0^\wedge = 1, \quad 1^\wedge = 2, \quad 2^\wedge = 0; \qquad 0^\vee = 2, \quad 1^\vee = 0, \quad 2^\vee = 1$$

By definition of $\times'$, etc., we have

$$(7.4)\qquad a \times' b = (a^{\wedge} \times b^{\wedge})^{\vee};\ a \times'' b = (a^{\vee} \times b^{\vee})^{\wedge};\ a +' b = (a^{\wedge} + b^{\wedge})^{\vee};$$
$$a +'' b = (a^{\vee} + b^{\vee})^{\wedge};\ a -' b = (a^{\wedge} - b^{\wedge})^{\vee};\ a -'' b = (a^{\vee} - b^{\vee})^{\wedge}.$$

In each of the relations (7.4) the $+$, $\times$ coordinate system is exhibited in a preferred role, in that each operation $\times', \times'', +', \cdots$ etc. is expressed in this coordinate system. This preferred role is removed, and each operation may then be expressed in any of the (three) permissible coordinate systems, by applying the *tri-ality* theorem, that is the p-ality theorem for $p = 3$. Each relation is then one of a *tri-al* set. From the first two relations (7.4), by tri-alization, we get for the tri-al ring products of a 3-ring, the

THEOREM 11. *Transformation Theorem* (= *'De Morgan' Formula for* 3 *Rings*). *In any 3-ring,*

$$(7.5)\qquad \begin{aligned} a \times' b &= (a^{\wedge} \times b^{\wedge})^{\vee} = (a^{\vee} \times'' b^{\vee})^{\wedge} \\ a \times'' b &= (a^{\vee} \times b^{\vee})^{\wedge} = (a^{\wedge} \times' b^{\wedge})^{\vee} \\ a \times b &= (a^{\wedge} \times'' b^{\wedge})^{\vee} = (a^{\vee} \times' b^{\vee})^{\wedge}. \end{aligned}$$

(It may be noted that the *simple* ring 'De-Morgan' theorem (1.5) may be obtained by degeneration from (7.5) by taking $a^{\wedge} = a^{\vee} = a^*$ and writing $\times' = \otimes$).

From its derivation it is easily seen that Theorem 11 gives the correct formulas not only for converting the ring products from one coordinate system to another, but also for converting any multitation (= operation) $\phi(a, b, \cdots)$ in the 3-ring. Thus, if ϕ', ϕ'' are the transforms (3.5) of ϕ by $^{\wedge}$ and by $^{\vee}$ respectively, then, as in (7.5), we have

$$(7.6)\qquad \begin{aligned} \phi'(a, b, \cdots) &= \{\phi(a^{\wedge}, b^{\wedge}, \cdots)\}^{\vee} = \{\phi''(a^{\vee}, b^{\vee}, \cdots)\}^{\wedge} \\ \phi''(a, b, \cdots) &= \{\phi(a^{\vee}, b^{\vee}, \cdots)\}^{\wedge} = \{\phi'(a^{\wedge}, b^{\wedge}, \cdots)\}^{\vee} \\ \phi(a, b, \cdots) &= \{\phi''(a^{\wedge}, b^{\wedge}, \cdots)\}^{\vee} = \{\phi'(a^{\vee}, b^{\vee}, \cdots)\}^{\wedge}. \end{aligned}$$

In the particular case of bitations we also write

$$\phi(a, b) = a\,\phi\,b$$

and the formulas (7.6) correspondingly, i.e.,

$$(7.7)\qquad \begin{aligned} a\,\phi'\,b &= (a^{\wedge}\,\phi\,b^{\wedge})^{\vee} = (a^{\vee}\,\phi''\,b^{\vee})^{\wedge} \\ a\,\phi''\,b &= (a^{\vee}\,\phi\,b^{\vee})^{\wedge} = (a^{\wedge}\,\phi'\,b^{\wedge})^{\vee} \\ a\,\phi\,b &= (a^{\wedge}\,\phi''\,b^{\wedge})^{\vee} = (a^{\vee}\,\phi'\,b^{\vee})^{\wedge}. \end{aligned}$$

The six general transformation formulas (7.6), and hence in particular (7.7) and (7.4), may be condensed into a single convenient formula by means of a simple rule of thumb. Let us agree to call the 'cyclical negation' operation, $^\wedge$, the *predecessor*, and $^\vee$ $(={}^{\wedge\wedge})$ the *successor* negation. Similarly, if k and κ are any two different members of the set $0, 1, 2$ we say that 'k is the *predecessor of* κ,'—also read, 'κ is the *successor* of k,' if $k \rightarrow \kappa$ in the 'standard' cyclic permutation (012). Since we have to do with a class of only three elements, we have the evident dichotomy: for any given k, κ with $k \neq \kappa$, of the two relations (1°): k is the predecessor of κ, (2°): κ is the predecessor of k, one and only one holds.

To *pre* a ring element, a, is to replace it by $a^\wedge$; to *pre* a ring operation is to replace it by its predecessor. A similar terminology holds with respect to *suc*. For instance, pre $\phi = \phi''$, suc $+ = +'$, suc $a = a + 2, \cdots$ etc. We may now reformulate the generalized Theorem 11, which we do in two slightly different ways (A) and $(\mathcal{A})$,—the latter for more convenient application.

THEOREM 12. FUNDAMENTAL TRANSFORMATION THEOREM FOR MULTITATIONS. *In a 3-ring, let* ϕ' *and* ϕ'' *be the transposes* (7.6) *of a multitation* ϕ *by* $^\wedge$ *and by* $^\vee$ *respectively; let* $\phi^{(k)}$, $\phi^{(\kappa)}$ *be elements of the set* ϕ, ϕ', ϕ''; *let* $^\Phi$, $^\Theta$ *be some arrangement of the set* $^\wedge$, $^\vee$.

(A) *For given* k *and* κ, *with* $k \neq \kappa$, *the formula expressing* $\phi^{(k)}$ *in terms of* $\phi^{(\kappa)}$ *is given by*

$$\phi^{(k)}(a, b, \cdots) = \{\phi^{(\kappa)}(a^\Phi, b^\Phi, \cdots)\}^\Theta, \tag{7.8}$$

where $^\Phi$ *must be chosen to 'agree with'* κ, *that is,* $^\Phi$ *is the predecessor negation,* $^\wedge$, *if* κ *is the predecessor of* k, *and is the successor negation,* $^\vee$, *if* κ *is the successor of* k.

$(\mathcal{A})$ *For given* k *and given* $^\Phi$,

$$\{\phi^{(k)}(a, b, \cdots)\}^\Phi = \phi^{(\kappa)}(a^\Phi, b^\Phi, \cdots), \tag{7.9}$$

where κ *must be chosen to 'agree with'* $^\Phi$, *in the above sense.*

By repeated application of $(\mathcal{A})$ of Theorem 12, and by use of the *pre*-ing and *suc*-ing terminology preceding Theorem 12, we have the very useful extension,

THEOREM 13. *Let* Ψ *be a multitation built up (by composition) from one or more 'component' multitations. Then a formula for* $\Psi^\wedge$ *is obtained* *by pre-ing everything in* Ψ, *that is, component multitations as well as ring element arguments. Similarly a formula for* $\Psi^\vee$ *is obtained by suc-ing everything in* Ψ.

As illustrations of the application of Theorem 13, we have

$$\begin{aligned} &\{(a\times b)\times'' c\}^\wedge = (a^\wedge\times'' b^\wedge)\times' c^\wedge \\ &\{(2+'a)\times''(b^\vee\times' c)\}^\vee = (0+''a^\vee)\times(b^\wedge\times'' c^\vee). \end{aligned} \tag{7.10}$$

Note. The treatment of 'constants' in Theorem 13 is exactly like that of a variable, unlike the 'contragredient' distinction required in the p-ality theorem.

Each of the relations (7. 5), as well as future identities, may be directly verified by expressing each side in the same coordinate system by means of (7. 1), (7. 2) and the relations (6. 4)-(6. 6) written for $p=3$, namely

$$\begin{aligned} &a\times' b = a\times b + a + b, && a\times'' b = a\times b + 2(a+b) + 2, \\ &a+' b = a + b + 1, && a+'' b = a + b + 2, \\ &a-' b = a - b - 1, && a-'' b = a - b - 2. \end{aligned} \tag{7.11}$$

Here again these exhibit the $+$, $\times$ coordinate system in a preferred role; this may be removed by tri-alizing the relations (7. 11), whence each operation, in any coordinate system, is expressible in terms of operations belonging entirely to any given coordinate system:

THEOREM 14. *In a 3-ring,*

$$\begin{aligned} a\times' b &= a\times b + a + b = a\times'' b +'' 2(a+'' b) +'' 0 \\ a\times'' b &= a\times b + 2(a+b) + 2 = a\times' b +' a +' b \\ a\times b &= a\times'' b +'' a +'' b = a\times' b +' 2(a+' b) +' 1 \\ a+' b &= a + b + 1 = a +'' b +'' 0 \\ a+'' b &= a + b + 2 = a +' b +' 0 \\ a + b &= a +'' b +'' 2 = a +' b +' 1 \\ a -' b &= a - b - 1 = a -'' b -'' 0 \\ a -'' b &= a - b - 2 = a -' b -' 0 \\ a - b &= a -' b -' 1 = a -'' b -'' 2. \end{aligned} \tag{7.12}$$

8. 3-ring-logics (continued). We now show that each 3-ring $(S, +, \times)$ is N-logically fixed, and moreover equationally. By the logic (= logical algebra) of a 3-ring we shall here always understand the N-logic,

$$(S, \times, \times', \times'', {}^\wedge, {}^\vee). \tag{8.1}$$

THEOREM 15. *Each 3-ring $(S, +, \times)$ is a ring-logic, with $+$ logically definable by the equation*

$$a + b = ab^\wedge \times' a^\wedge b \times' a^2 b^2. \tag{8.2}$$

Proof. The identity (8.2) may be verified by direct substitution from (7.1) and (7.12), making use of the 3-ring definitive properties (6.1) and (6.2) for $p = 3$. In the $+$, $\times$ coordinate system the right of (8.2) then becomes

$$ab^\wedge + a^\wedge b + ab^\wedge a^\wedge b + a^2b^2 + a^2b^2(ab^\wedge + a^\wedge b + ab^\wedge a^\wedge b), \tag{8.3}$$

which readily reduces to $a + b$. There remains to show that $(S, +, \times)$ is fixed by its logic. Suppose $(S, +_1, \times)$ is a ring having the same logic as $(S, +, \times)$. We must show that $+ = +_1$. By hypothesis

$$1 + a = 1 +_1 a, \tag{8.4}$$

from which one finds that $3a = 0$. Hence, since $\times$ is the same for both rings, $(S, +_1, \times)$ is also a 3-ring, by definition (6.1), (6.2). We may now re-verify (8.2) with $a +_1 b$ on the left, which shows that $+ = +_1$, and proves Theorem 15.

By tri-alizing (8.2), either directly as given or after expressing the right side in various 'pure' or 'mixed' ways by use of the transforming relations (7.12), one may obtain many similar formulas. We here note only one example. If we start with (8.2) expressed in a pure form, we get the tri-al set:

THEOREM 16. *In a 3-ring,*

$$a + b = \{(ab^\wedge)^\wedge (a^\wedge b)^\wedge (a^2b^2)^\wedge\}^\vee \tag{8.5}$$

$$a +' b = \{(a \times' b^\wedge)^\wedge \times' (a^\wedge \times' b)^\wedge \times' (a \times' a \times' b \times' b)^\wedge\}^\vee \tag{8.6}$$

$$a +'' b = \{(a \times'' b^\wedge)^\wedge \times'' (a^\wedge \times'' b)^\wedge \times'' (a \times'' a \times'' b \times'' b)^\wedge\}^\vee. \tag{8.7}$$

Again, from each formula for $a +' b$, such as (8.6), and similarly for $a +'' b$, by recalling that $(a +' b)^\vee = a + b$ and $(a +'' b)^\wedge = a + b$, and by use of Theorem 13, we obtain numerous new formulas for $a + b$; we here mention one such, obtained from (8.6),

$$a + b = (a^\vee \times'' b)(a \times'' b^\vee)(a^\vee \times'' a^\vee \times'' b^\vee \times'' b^\vee). \tag{8.8}$$

9. 3-valued logic. We here give a very brief orientation of 3-valued logic within the framework of general 3-ring-logics, and consider an illustration of the tri-ality theorem applied to the former. A more comprehensive treatment of 3-valued-logic, with the general p-ality theory as background, will be offered in a later communication.

Exactly as the logic of propositions ($=$ 2-valued logic) is mathematically equivalent to the simplest 2-ring ($=$ Boolean ring) $F_2 =$ ring ($=$ field) of 2 elements or 'truth values' 0, 1, so is the 3-valued logic equivalent to the simplest 3-ring $F_3 =$ ring ($=$ field) of 3 elements or 'truth values' 0, 1, 2.

By a well known theorem (holding as well in $F_p =$ field of residues mod $p =$ prime), (I): each multitation $\phi(x, y, \cdots)$ of the set F_3 may be 'analytically' expressed,—and moreover, uniquely, as a polynomial, mod 3, of the field $(F_3, +, \times)$. Thus the 3^3 monotations of the set F_3 are uniquely 'analytically' exhibited in the ring language by

$$\phi(x) = a + bx + cx^2, \tag{9.1}$$

and the $3^{(3^2)}$ bitations uniquely by

$$\phi(x, y) = a + bx + cy + dxy + ex^2 + fy^2 + gxy^2 + hx^2y + kx^2y^2; \tag{9.2}$$

etc.

The 'logical language' is concerned entirely with

$$(F_3, \times, \times', \times'', {}^\wedge, {}^\vee), \tag{9.3}$$

and the 'completeness' of this logic follows from (I) and Theorem 15, making it possible to formulate all possible multitations of F_3, i. e., all possible 3-valued propositional functions, entirely within the logic (9.3).

Moreover, since in F_3,

$$| \times, {}^\wedge | = | \times, {}^\vee | = | \times', {}^\wedge | = \cdots = | \times, \times', \times'', {}^\wedge, {}^\vee |, \tag{9.4}$$

(see (4.7) and (4.8)), in a formal development of this 3-valued logic of propositions it is sufficient to take only the operations $\times, {}^\wedge$, or else only $\times, {}^\vee$, etc., as undefined.

We consider an illustration of tri-al propositions. Let us read

$$0 = \text{false}, \qquad 1 = \text{true}, \qquad 2 = \text{indeterminate}. \tag{9.5}$$

Let $a \& b$ denote the propositional function which is true only if a and b are both true, and false otherwise. In the ring language (9.2) we find (uniquely),

$$a \mathbin{\&} b = ab + ab^2 + a^2b + a^2b^2, \tag{9.6}$$

and in the logical language

$$a \mathbin{\&} b = ab(a \times' b)^{\wedge} = aa^{\wedge}bb^{\wedge}. \tag{9.7}$$

The tri-als of (9. 7) are

$$\begin{aligned} a \mathbin{\&'} b &= a \times' b \times' (a \times'' b) = a \times' a^{\wedge} \times' b \times' b^{\wedge} \\ a \mathbin{\&''} b &= a \times'' b \times'' (ab)^{\wedge} = a \times'' a^{\wedge} \times'' b \times'' b^{\wedge}. \end{aligned} \tag{9.8}$$

Here, in words,

(9. 9)
$a \mathbin{\&'} b$ is false only if a and b are both false, and indeterminate otherwise.
$a \mathbin{\&''} b$ is indeterminate only if a and b are both true, and true otherwise.

10. Conjectures, problems. We have seen that the cyclic negation group N is fully adapted to both 2-rings (Boolean rings), and 3-rings; otherwise stated, that a p-ring is a ring-logic (N) for $p = 2$ and $p = 3$. Is this true for all primes p, or indeed for *some* prime $p > 3$? This question still remains unanswered. It may be shown, exactly as for 3-rings, that a p-ring is N-logically fixed if it is N-logically equationally definable. Hence the above question has an affirmative answer for such and only such primes p for which an identity similar to (8. 2) exists.

In the case of a Boolean ring (2-ring) the group N reduces to the complementation group C, and hence in a Boolean ring $* = {}^{\wedge}$, $\otimes = \times'$, $\oplus = +'$, etc. (see **1**). It is instructive to compare the logical formulas for $+$ given in (1. 8) and (8. 2) for $p = 2$ and for $p = 3$,

$$a + b = ab^{\wedge} \times' a^{\wedge}b \quad \text{(in a 2-ring)} \tag{10.1}$$

$$a + b = ab^{\wedge} \times' a^{\wedge}b \times' a^2b^2 \quad \text{(in a 3-ring).} \tag{10.2}$$

It is easily checked that the 3-case does not 'cover' the 2-case, i. e., that the formula (10. 2) does not give the correct definition for $+$ when it is applied to a 2-ring; similarly the 2-case does not cover the 3-case. There are of course other logical equational formulas for $+$, such as (8. 5) and (8. 8), and others,—both for 2- and for 3-rings. It is however to be conjectured that there exists no single formula which covers both cases, in the above sense; and similarly for any primes p, p' $(p' \neq p)$ for which p-ring and p'-ring are both ring logics.

In this connection the formulas (10.1) and (10.2) suggest that a similar formula with 5 factors might exist for 5-rings, of the form

$$a + b = ab^{\wedge} \times' a^{\wedge}b \times' W \times' Y \times' Z. \tag{10.3}$$

It may, however, be shown that this is impossible, in fact that no formula of the type (10.3) exists which contains a 'factor' $ab^{\wedge} \times' a^{\wedge}b$.[9]

UNIVERSITY OF CALIFORNIA,
BERKELEY.

BIBLIOGRAPHY.

1. A. L. Foster, "The theory of Boolean-like rings," *Transactions of the American Mathematical Society*, vol. 59 (1946), pp. 166-187.
2. ———, "The idempotent elements of a commutative ring form a Boolean algebra; ring duality and transformation theory," *Duke Mathematical Journal*, vol. 12 (1945), pp. 143-152.
3. ———, "Maximal idempotent sets in a ring with unit," *Duke Mathematical Journal*, vol. 13 (1946), pp. 247-258.
4. ———, "On the permutational representation of general sets of operations by partition lattices," *Transactions of the American Mathematical Society*, vol. 66 (1949), pp. 366-388.
5. A. L. Foster and B. A. Bernstein, "Symmetric approach to commutative rings with duality theorem: Boolean duality as special case," *Duke Mathematical Journal*, vol. 11 (1944), pp. 603-616.
6. ———, "A dual symmetric definition of field," *American Journal of Mathematics*, vol. 67 (1945), pp. 329-349.
7. F. Harary, "The structure of Boolean-like rings," in process of publication.
8. M. H. Stone, "The theory of representations of Boolean algebras," *Transactions of the American Mathematical Society*, vol. 40 (1936), pp. 37-111.
9. N. H. McCoy and Deane Montgomery, "A representation of generalized Boolean rings," *Duke Mathematical Journal*, vol. 3 (1937), pp. 455-459.

---

[9] (Added in proof). Since this article was presented, the author has succeeded in confirming the above conjecture; it is shown that all p-rings are ring-logics (mod N). As might be expected for $p > 3$ the logical definition of $+$, corresponding to the special cases (10.1) and (10.2), is quite complicated. This result, which is to appear in the *University of California Publications in Mathematics*, is based on a comprehensive study of the structure of p-rings, in process of publication in *Acta Mathematica*.

ON LINEAR DIFFERENCE EQUATIONS OF SECOND ORDER.*

By PHILIP HARTMAN and AUREL WINTNER.

The theorems to be proved below represent extensions to the case of difference equations of certain results proved in [2] for the case of differential equations. The standard proof of the theorem of Kneser, used *loc. cit.*, is not now available and will have to be replaced by another approach. The latter will be patterned after the method of [3]. It turns out that the resulting criterion is by necessity different from that prevailing in the case of differential equations.

The first of the theorems to be proved is as follows:

(I) *Let* $q_0, q_1, \cdots; r_0, r_1, \cdots$ *be two sequences of real numbers satisfying the inequalities*

$$1 - r_k - q_k > 0 \text{ and } q_k > 0 \qquad (k = 0, 1, \cdots). \tag{1}$$

Then the difference equation

$$\Delta^2 y_k + r_k \Delta y_k - q_k y_k = 0 \qquad (k = 0, 1, \cdots) \tag{2}$$

possesses a solution $y_0, y_1, \cdots$ *satisfying*

$$y_k > 0 \text{ and } \Delta y_k > 0 \qquad (k = 0, 1, \cdots). \tag{3}$$

It is understood that $\Delta y_k = y_{k+1} - y_k$ and $\Delta^2 y_k = \Delta(\Delta y_k) = y_{k+2} - 2y_{k+1} + y_k$.

Kneser's theorem, which deals with the differential equation

$$y'' - q(x)y = 0, \tag{4}$$

was extended in [2] to the equation

$$y'' + r(x)y' - q(x)y = 0. \tag{5}$$

In the case of differential equations, it is supposed that $r(x)$, $q(x)$ are continuous for large x, and that $q(x) \geqq 0$, but no restriction is placed on $r(x)$. If $r(x) \equiv 0$, the analogue of conditions (1) becomes

$$1 - q(x) > 0 \text{ and } q(x) > 0.$$

---

* Received March 5, 1949.

The first of the latter two inequalities is not needed in the theorem on differential equations. On the other hand, (I) becomes false if the first condition of (1) is omitted. In fact, the constant 1 occurring in this condition is the best constant; in the sense that if the first inequality of (1) is replaced by $1-\epsilon-r_k-q_k>0$, where $\epsilon>0$, then (2) need not have a solution satisfying (3). This is illustrated by the equation $\Delta^2 y_k-(1+\epsilon)y_k=0$, where $k=0,1,2,\cdots$. In fact, every solution $y_0, y_1, \cdots$ of this equation is a linear combination of the solutions given by $y_k=(1\pm(1+\epsilon)^{\frac{1}{2}})^k$, where $k=0,1,\cdots$, but no linear combination can satisfy (3) when $\epsilon>0$.

If $1-r_k-q_k>0$ is weakened to $1-r_k-q_k\geqq 0$, then the assertion (3) must be weakened to allow $y_k=0$ for all large k. This is illustrated by case $\epsilon=0$ of the equation just considered; actually, the equation then reduces to the first order difference equation $y_{k+2}-2y_{k+1}=0$, where $k=0,1,\cdots$, which possesses $y_0=\text{const.}\geqq 0$, $y_1=y_2=\cdots=0$ as the only non-decreasing solution.

Corresponding to the situation in differential equations, the second condition in (1) can be relaxed to $q_k\geqq 0$, provided that q_k does not vanish for all large k. In the latter case, the second assertion in (2) must be relaxed to $\Delta y_k=0$ for large k.

It is known (Sturm; cf. [1], pp. 176-177) that the first inequality in (1) assures that the "zeros" of two non-trivial solutions of (2) separate each other. A "zero" is meant in the following sense: If $y_0, y_1, \cdots$ is a solution of (2), and if one considers the polygonal path joining the points (n, y_n) in the (x, y)-plane, then a point common to this path and to the x-axis is called a zero of the solution $y_0, y_1, \cdots$.

It will be shown that a non-trivial solution of (2) has at most one zero in virtue of (1). In view of the separation theorem, it is sufficient to exhibit a solution of (1) which has no zero. Such a solution is obtained by assigning, for example, the initial conditions $y_0=1$, $\Delta y_0=y_1-y_0=1$; the solution $y_0, y_1, \cdots$ is then determined uniquely, since (2) can be written in the form

$$y_{k+2}-(2p_k-r_k)y_{k+1}+(1-r_k-q_k)y_k=0 \qquad (k=0,1,\cdots).$$

That the solution, determined by the assigned initial conditions, has no zero is a consequence of the following assertion:

If $y_0, y_1, \cdots$ is a solution of (2) and if, for some fixed $n\geqq 0$,

$$y_n\geqq 0 \text{ and } \Delta y_n\geqq 0, \tag{6}$$

then

$$y_k\geqq y_n;\text{ in fact, } \Delta y_k\geqq 0 \qquad (k=n, n+1,\cdots). \tag{7}$$

In order to see this, rewrite the case $k=n$ of (2) in the form

$$\Delta y_{n+1}+(r_n-1)\Delta y_n=qy_n\geqq 0.$$

The first condition of (1) implies that $1-r_n>q_n\geqq 0$. Consequently, by the last formula line, $\Delta y_{n+1}\geqq 0$. Hence, (7) holds for $k=n+1$ and, by induction, for all $k\geqq n$.

Since a non-trivial solution of (2) has at most one zero, it follows that if n, m is a given pair of integers satisfying $0\leqq m<n$, and if y_m, y_n is a given pair of numbers, then there exists one and only one solution $y_0, y_1, \cdots$ for which y_n, y_m assume the given values. In fact, every solution of (2) is determined by its pair of initial values y_0, y_1; hence, every solution $y_0, y_1, \cdots$ is a linear combination,

$$y_k=c_1y^1{}_k+c_2y^2{}_k, \tag{8}$$

of the pair of solutions $y^1{}_0, y^1{}_1, \cdots, y^2{}_0, y^2{}_1, \cdots$ determined by $y^1{}_0=1$, $y^1{}_1=0$ and $y^2{}_0=0$, $y^2{}_1=1$, respectively. Conversely, every linear combination (8) is a solution of (2). Thus, what is required for the existence and the uniqueness of a solution, for which y_n, y_m are prescribed, is the unique solvability for c_1, c_2 of the linear equations

$$y_m=c_1y^1{}_m+c_2y^2{}_m \text{ and } y_m=c_1y^1{}_m+c_2y^2{}_m.$$

If the latter do not have a unique solution c_1, c_2, then there exists a pair of constants c_1, c_2, not both zero, such that

$$c_1y^1{}_m+c_2y^2{}_m=0 \text{ and } c_1y^1{}_n+c_2y^2{}_n=0.$$

But then the corresponding non-trivial solution (8) has two zeros, which is impossible. Consequently, if $m\neq n$, the numbers y_m, y_n determine a unique solution of (2).

In order to complete the proof of (I), let $y_0{}^j, y_1{}^j, y_2{}^j, \cdots$ denote the unique solution of (2) satisfying

$$y_0{}^j=1 \text{ and } y_j{}^j=0, \qquad (j=1,2,\cdots). \tag{9}$$

Then $y_k{}^j>0$ for $k=0,1,\cdots,j-1$, since otherwise the solution $y_0{}^j, y_1{}^j, \cdots$ had two zeros. Also

$$1=y_0{}^j>y_1{}^j>\cdots>y_j{}^j=0. \tag{10}$$

For, if $y_n{}^j\geqq y_{n+1}{}^j$, that is, if $\Delta y_n{}^j\geqq 0$ holds for some value of n on the interval $0\leqq n\leqq j-1$, then $y_k{}^j\geqq y_n{}^j>0$ for $k\geqq n$, since (6) implies (7). But this contradicts $y_j{}^j=0$. Consequently, (10) holds.

A diagonal selection process shows that the sequence of integers $j=1, 2, \cdots$, contains a subsequence having the property that, if the j-th element of the subsequence is denoted simply by j, then, as $j \to \infty$, the limit $y_k = \lim y_k{}^j$ exists for $k=0, 1, \cdots$. Clearly, this limit sequence $y_0, y_1, \cdots$ is a solution of (2) and satisfies

$$y_k \geqq 0 \text{ and } \Delta y_k \leqq 0, \qquad (k=0,1,\cdots) \tag{11}$$

in virtue of (10). Furthermore, $y_0 = 1$, by (9); so that the solution $y_0, y_1, \cdots$ is not identically zero. Since it has at most one zero, it follows, from (11), that it has no zero, that is, that $y_k > 0$.

It remains to show that $\Delta y_n = 0$ cannot hold for any n. If this did occur for some n, then $\Delta y_k = 0$ for $k \geqq n$, by (11), since (6) implies (7). Hence, $\Delta^2 y_k = 0$ for $k \geqq n$; so that $q_k y_k = 0$ for $k \geqq n$, by (2). This implies $q_k = 0$ for $k \geqq n$, which contradicts the second assumption in (1).

The proof of (I) is now complete.

(II) *Let* $p_0, p_1, \cdots$; $q_0, q_1, \cdots$; $r_0, r_1, \cdots$ *be three sequences of numbers satisfying*

$$p_k > 0;\ p_k - r_k - q_k > 0;\ q_k > 0, \qquad (k=0,1,\cdots), \tag{12}$$

and let the three sequences $\Delta p_0, \Delta p_1, \cdots$; $q_0, q_1, \cdots$; $r_0, r_1, \cdots$ *be completely monotone, that is, let*

$$(-1)^n \Delta^{n+1} p_k \geqq 0; \qquad (-1)^n \Delta^n q_k \geqq 0; \qquad (-1)^n \Delta^n r_k \geqq 0, \tag{13}$$

where $k, n = 0, 1, 2, \cdots$. *Then the difference equation*

$$p_k \Delta^2 y_k + r_k \Delta y_k - q_k y_k \qquad (k=0,1,\cdots) \tag{14}$$

possesses a positive, completely monotone solution $y_0, y_1, \cdots$:

$$y_k > 0 \text{ and } (-1)^n \Delta^n y_k \geqq 0 \qquad (k, n = 0, 1, \cdots). \tag{15}$$

It is understood that

$$\Delta^n y_k = \Delta(\Delta^{n-1} y_k) = \sum_{m=0}^{n} C_m{}^n (-1)^{n-m} y_{m+k},$$

where the $C_m{}^n$ denote the binomial coefficients. The theorem (II) is an analogue of a theorem on differential equations proved in [2].

The proof of (II) proceeds as follows: If (14) is divided by p_k, which is permissible in view of the first assumption in (12), then the difference equation (14) is reduced to one of the form (2). Furthermore, the last two conditions of (12) imply the conditions (1). Hence, (14) possesses a solution $y_0, y_1, \cdots$ satisfying (3). This means that

$$(16_n) \qquad (-1)^n \Delta^n y_k \geqq 0 \qquad (k = 0, 1, \cdots)$$

holds for $n = 0, 1$. In order to prove that (16_n) holds for every n, suppose that it holds for $n = 0, 1, \cdots, j+1$.

If $j \geqq 0$ is fixed and if $a_0, a_1, \cdots; b_0, b_1, \cdots$ are given sequences, then

$$\Delta^j (a_k b_k) = \sum_{m=0}^{j} C_m{}^j (\Delta^{j-m} b_{k+m}) (\Delta^m a_k), \qquad (k = 0, 1, \cdots).$$

Thus, if the operator Δ^j is applied to (14) and the resulting equation is solved for $\Delta^{j+2} y_k$, it is seen that $p_{k+j} \Delta^{j+2} y_k$ equals

$$-\sum_{m=0}^{j-1} C_m{}^j (\Delta^{j-m} p_{k+m})(\Delta^{m+2} y_k) - \sum_{m=0}^{j} C_m{}^j (\Delta^{j-m} r_{k+m})(\Delta^{m+1} y_k) + \sum_{m=0}^{j} C_m{}^j (\Delta^{j-m} q_{k+m})(\Delta^m y_k).$$

The induction hypothesis and (13) show that every term in these sums vanishes or has the same sign as $(-1)^{j+2}$; hence,

$$(-1)^{j+2} p_{k+j} \Delta^{j+2} y_k \geqq 0 \qquad (k = 0, 1, \cdots).$$

Since $p_{k+j} > 0$, the induction and the proof of (II) are now complete.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] M. Bôcher, "Boundary problems in one dimension," *5th International Congress of Mathematicians, Proceedings*, vol. I (1912), pp. 163-195.

[2] P. Hartman and A. Wintner, "On the Laplace-Fourier transcendents," *American Journal of Mathematics*, vol. 71 (1949), pp. 367-372.

[3] A. Wintner, "On linear repulsive forces," *ibid.*, pp. 362-366.

ON THE UNIFORM CESÀRO SUMMABILITY OF CERTAIN SPECIAL TRIGONOMETRICAL SERIES.*

By CHING-TSÜN LOO.

1. It is well-known that if $\lambda_\nu \to 0, \Delta\lambda_\nu \geq 0$,[1] then both series

$$(1.1) \qquad \tfrac{1}{2}\lambda_0 + \sum_{\nu=1}^{\infty} \lambda_\nu \cos \nu\theta, \qquad \sum_{\nu=1}^{\infty} \lambda_\nu \sin \nu\theta$$

are uniformly convergent in $(\epsilon, \pi - \epsilon)$, $\epsilon > 0$. We also know that if $\lambda_\nu \to 0$, and $\Delta^2\lambda_\nu \geq 0$, then the first derived series of (1.1) are uniformly summable $(C, 1)$ in that interval.[2] It is interesting to see whether these theorems can be extended to a theorem of general scale. The purpose of this paper is to give a positive answer to this question.

THEOREM. *If* $\lambda_\nu \to 0$ *and* $\Delta^{k+1}\lambda_\nu \geq 0$, *then the k-th derived series of* (1.1) *are uniformly summable* (C, k) *in any interval* $(\epsilon, \pi - \epsilon)$, $\epsilon > 0$, *where k is any integer* ≥ 0.

We write

$$(1.2) \qquad C_\nu^{(k)} = \lambda_\nu h_\nu^{(k)}, \qquad h_\nu^{(k)} = \nu^k A_{n-\nu}^{(k)}$$

where $A_\nu^{(k)} = \binom{k+\nu}{\nu}$ are the Cesàro numbers. We have to prove that

$$(1.3) \qquad \sigma_n^{(k)}(\theta) = \sum_{\nu=0}^{n} C_\nu^{(k)} e^{i\nu\theta} / A_n^{(k)}$$

is uniformly convergent in the interval $(\epsilon, \pi - \epsilon)$ as $n \to \infty$. Our theorem then follows by considering the semi-sums and semi-differences of $\sigma_n^{(k)}(\theta)$ and $\sigma_n^{(k)}(-\theta)$.

2. We shall first establish the following formula: If $p \geq 1$, then

$$\begin{aligned}(2.1) \quad A_n^{(k)}\sigma_n^{(k)}(\theta) = & \sum_{j=0}^{p-1} (-1)^j e^{ij\theta} \Delta^j C_0^{(k)} / (1 - e^{i\theta})^{j+1} \\ & + (-1)^p e^{ip\theta} / (1 - e^{i\theta})^p \sum_{\nu=0}^{n-p} e^{i\nu\theta} \Delta^p C_\nu^{(k)} \\ & + e^{i(n+1)\theta} \sum_{j=0}^{p-1} (-1)^{j+1} / (1 - e^{i\theta})^{j+1} \Delta^j C_{n-j}^{(k)}.\end{aligned}$$

---

* Received March 21, 1949.

[1] We write $\Delta^0\lambda_\nu = \lambda_\nu$, $\Delta^1\lambda_\nu = \Delta\lambda_\nu = \lambda_\nu - \lambda_{\nu+1}$ and $\Delta^k\lambda_\nu = \Delta(\Delta^{k-1}\lambda_\nu)$ for $k \geq 1$.

[2] A. Zygmund, *Trigonometrical series* (Warszawa-Lwów, 1935), p. 129, Ex. 6.

Let

$$S_\nu = 1 + e^{i\theta} + \cdots + e^{i\nu\theta} = (1 - e^{i(n+1)\theta})/(1 - e^{i\theta}),$$

then

$$S_\nu - S_{\nu-1} = -\,(e^{i(\nu+1)\theta} - e^{i\nu\theta})\,/(1 - e^{i\theta})$$

and

$$(2.2)\qquad \sum_{\nu=0}^{n} C_\nu^{(k)} e^{i\nu\theta} = \sum_{\nu=0}^{n} C_\nu^{(k)} (S_\nu - S_{\nu-1})$$

$$= -\sum_{\nu=0}^{n} C_\nu^{(k)} (e^{i(\nu+1)\theta} - e^{i\nu\theta})/(1 - e^{i\theta})$$

$$= \{C_0^{(k)} - e^{i\theta}\sum_{\nu=0}^{n-1} e^{i\nu\theta}\Delta C_\nu^{(k)} - e^{i(n+1)\theta}C_n^{(k)}\}\,/(1 - e^{i\theta}),$$

where $C_0^{(k)} = 0$. Hence the formula (2.1) has been proved for $p = 1$.

The same device gives

$$(2.3)\qquad \sum_{\nu=0}^{n-(p-1)} e^{i\nu\theta}\Delta^{(p-1)}C_\nu^{(k)} = \{\Delta^{(p-1)}C_0^{(k)} - e^{i\theta}\sum_{\nu=0}^{n-p} e^{i\nu\theta}\Delta^{p}C_\nu^{(k)}$$

$$- e^{i(n-p+2)}\Delta^{(p-1)}C_{n-p+1}^{(k)}\}/(1 - e^{i\theta}).$$

Suppose (2.1) holds for $p - 1$. Replacing the middle term of the right side of (2.1) by left side of (2.3) and rearranging the corresponding terms, we get (2.1) for p.

3. Let us put $k + 1$ for p in (2.1). Let $S_{n\,1}(\theta)$, $S_{n\,2}(\theta)$ and $S_{n\,3}(\theta)$ be the first, second and third sums of the right side of that formula. We are going to prove that $S_{n\,3}(\theta)/A_n^{(k)} = o(1)$ and that each of the sums $S_{n\,1}(\theta)/A_n^{(k)}$ and $S_{n\,2}(\theta)/A_n^{(k)}$ tends to a finite limit uniformly with respect to θ in $(\epsilon, \pi - \epsilon)$.

In order to estimate the orders of $\Delta^l C_\nu^{(k)}$ for $0 \leq l \leq k + 1$, we use the formula

$$(3.1)\qquad \Delta^l \xi_\nu \eta_\nu = \sum_{q=0}^{l} C_{l\,q}\Delta^q\xi_\nu\Delta^{l-q}\eta_{\nu+q},$$

where C_{lq} are constants. First of all we have

$$(3.2)\qquad \Delta^l h_\nu^{(k)} = \Delta^l \nu^k A_{n-\nu}^{(k)} = \sum_{q=0}^{l} C_{l\,q}\Delta^q A_{n-\nu}^{(k)}\Delta^{l-q}(\nu + q)^k$$

$$= \sum_{q=0}^{l} C_{l\,q} A_{n-\nu}^{(k-q)}\Delta^{l-q}(\nu + q)^k,$$

since $\Delta A_{n-\nu}^{(k)} = A_{n-\nu}^{(k)} - A_{n-(\nu+1)}^{(k)} = A_{n-\nu}^{(k-1)}$, and in general $\Delta^l A_{n-\nu}^{(k)} = A_{n-\nu}^{(k-l)}$ for $0 \leq l \leq k$, and is identically zero for $l > k$. Next we use (3.1) again,

$$(3.3)\qquad \Delta^l C_\nu{}^{(k)} = \Delta^l \lambda_\nu h_\nu{}^{(k)} = \sum_{p=0}^{l} C_{l\ p} \Delta^p h_\nu{}^{(k)} \Delta^{l-p} \lambda_{\nu+p}$$
$$= \sum_{p=0}^{l} C_{l\ p} \sum_{q=0}^{p} C_{p\ q} A_{n-\nu}{}^{(k-q)} \Delta^{p-q} (\nu+q)^k \Delta^{l-p} \lambda_{\nu+p}.$$

Observe that the terms in $S_{n\,3}(\theta)/A_n{}^{(k)}$ apart from corresponding factors $(-1)^{l+1} e^{i(n+1)\theta}/(1-e^{i\theta})^{l+1}$ are each of the form $\Delta^l C_{n-l}{}^{(k)}/A_n{}^{(k)}$, $l = 0, 1, \cdots, k$. Using (3.3), we obtain

$$\Delta^l C_{n-l}{}^{(k)}/A_n{}^{(k)} = O\Big(\sum_{p=0}^{l} \sum_{q=0}^{p} n^{k-p+q} \Delta^{l-p} \lambda_{n-l+p}/n^k\Big)$$
$$= O\Big(\sum_{p=0}^{l} \sum_{q=0}^{p} n^{-(p-q)} \Delta^{(l-p)} \lambda_{n-l+p}\Big) = o(1),$$

since $l \geq p \geq q$, and $\Delta^{l-p} \lambda_{n-l+p} = o(1)$. Thus we have

$$(3.4)\qquad S_{n\,3}/A_n{}^{(k)} = e^{i(n+1)\theta}/A_n{}^{(k)} \sum_{j=0}^{k} (-1)^{j+1}/(1-e^{i\theta})^{j+1} \Delta^j C_{n-j}{}^{(k)} = o(1)$$

uniformly in $(\epsilon, \pi - \epsilon)$.

Next observe that terms in $S_{n\,1}(\theta)/A_n{}^{(k)}$ apart from corresponding factors $(-1)^l e^{il\theta}/(1-e^{i\theta})$ are each of the form $\Delta^l C_\nu{}^{(k)}/A_n{}^{(k)}$, $l = 0, 1, \cdots, k$. Since we have

$$\Delta^l C_0{}^{(k)}/A_n{}^{(k)} = 1/A_n{}^{(k)} \sum_{q=0}^{l} (-1)^q \binom{l}{q} A_{n-q}{}^{(k)} q^k \lambda_q$$
$$= \sum_{q=0}^{l} (-1)^q \binom{l}{q} A_{n-q}{}^{(k)} q^k \lambda_q / A_n{}^{(k)}$$

which tends to $\sum_{q=0}^{l} (-1)^q \binom{l}{q} q^k \lambda_q$ as $n \to \infty$, it follows that

$$(3.5)\qquad S_{n\,1}(\theta)/A_n{}^{(k)} = 1/A_n{}^{(k)} \sum_{j=0}^{k} (-1)^j e^{ij\theta} \Delta^j C_0{}^{(k)}/(1-e^{i\theta})^{j+1}$$
$$\to \sum_{j=0}^{k} \sum_{q=0}^{j} (-1)^{j+q} \binom{j}{q} e^{ij\theta} q^k \lambda_q/(1-e^{i\theta})^{j+1}.$$

4. It remains to prove that

$$(4.1)\qquad S_{n\,2}(\theta)/A_n{}^{(k)} = \{(-1)^{k+1} e^{i(k+1)\theta}/(1-e^{i\theta})^{k+1} A_n{}^{(k)}\} \sum_{\nu=0}^{n-(k+1)} e^{i\nu\theta} \Delta^{k+1} C_\nu{}^{(k)}$$

tends uniformly to a limit.

To make the situation clear, we shall separate $\Delta^{k+1} C_\nu{}^{(k)}/A_n{}^{(k)}$ into three sums. By (3.2),

$$\Delta^{k+1} h_\nu{}^{(k)} = \sum_{q=1}^{k} C_{k+1\ q} A_{n-\nu}{}^{(k-q)} \Delta^{k+1-q} (\nu+q)^k$$

since $\Delta^{k+1}A_{n-\nu}^{(k)}=0$, $\Delta^{k+1}(\nu+q)^k=0$. By (3. 3),

$$\begin{aligned}\Delta^{k+1}C_\nu^{(k)}/A_n^{(k)} &= 1/A_n^{(k)}\sum_{p=0}^{k}C_{k+1\ p}\sum_{q=0}^{p}C_{p\ q}A_{n-\nu}^{(k-q)}\Delta^{p-q}(\nu+q)^k\Delta^{k+1-p}\lambda_{\nu+p}\\ &\quad+C_{k+1\ k+1}/A_n^{(k)}\sum_{q=1}^{k}C_{k+1\ q}A_{n-\nu}^{(k-q)}\Delta^{k+1-q}(\nu+q)^k\lambda_{\nu+k+1}\\ &= 1/A_n^{(k)}\sum_{p=0}^{k}C_{k+1\ p}C_{p\ 0}A_{n-\nu}^{(k)}\Delta^{p}\nu^k\Delta^{k+1-p}\lambda_{\nu+p}\\ &\quad+1/A_n^{(k)}\sum_{p=1}^{k}C_{k+1\ p}\sum_{q=1}^{p}C_{p\ q}A_{n-\nu}^{(k-q)}\Delta^{p-q}(\nu+q)^k\Delta^{k+1-p}\lambda_{\nu+p}\\ &\quad+C_{k+1\ k+1}/A_n^{(k)}\sum_{q=1}^{k}C_{k+1\ q}A_{n-\nu}^{(k-q)}\Delta^{k+1-q}(\nu+q)^k\lambda_{\nu+k+1}\end{aligned}$$

$=J_{n\ \nu\ 1}+J_{n\ \nu\ 2}+J_{n\ \nu\ 3}$, say. Putting

$$(4.\ 2)\quad 1/A_n^{(k)}\sum_{\nu=0}^{n-(k+1)}e^{i\nu\theta}\Delta^{k+1}C_\nu^{(k)}=\sum_{\nu=0}^{n-(k+1)}J_{n\ \nu\ 1}e^{i\nu\theta}+\sum_{\nu=0}^{n-(k+1)}J_{n\ \nu\ 2}e^{i\nu\theta}+\sum_{\nu=0}^{n-(k+1)}J_{n\ \nu\ 3}e^{i\nu\theta}$$

$=W_{n\ 1}(\theta)+W_{n\ 2}(\theta)+W_{n\ 3}(\theta)$, we are going to prove that $W_{n\ 2}(\theta)=o(1)$, $W_{n\ 3}(\theta)=o(1)$ uniformly in θ, and that $W_{n\ 1}(\theta)$ tends uniformly to a limit. Before doing so, we shall first deduce from our assumptions $\lambda_\nu\to 0$, $\Delta^{k+1}\lambda_\nu\geqq 0$, some simple consequences which are useful in the following proofs.

We observe that the assumptions imply that

$$(4.\ 3)\quad \begin{array}{ll}\text{(i)}\quad \Delta^l\lambda_\nu\geqq 0, & \text{(ii)}\quad \Delta^l\lambda_\nu=o(\nu^{-l}),\\ \text{(iii)}\quad \sum_{\nu=0}^{\infty}(\nu+1)^l\Delta^{l+1}\lambda_\nu<\infty, & \text{(iv)}\quad \sum_{\nu=0}^{n}(\nu+1)^{l+s}\Delta^{l+1}\lambda_\nu=o(n^s),\\ & \qquad s>0,\end{array}$$

for any $l=0,1,2,\cdots,k$. Since $\Delta^{k+1}\lambda_\nu\geqq 0$, that is $\Delta^k\lambda_\nu\geqq\Delta^k\lambda_{\nu+1}$, the sequence $\{\Delta^k\lambda_\nu\}$ decreases to zero. Hence $\Delta^k\lambda_\nu\geqq 0$. This implies $\Delta^{k-1}\lambda_\nu\geqq 0$ and so on. Finally we get $\Delta\lambda_\nu\geqq 0$. Since $\lambda_\nu\to 0$, $\Delta\lambda_\nu\geqq 0$, we have

$$(4.\ 4)\qquad \sum_{\nu=0}^{\infty}\Delta\lambda_\nu=\lambda_0,$$

from which, on account of the fact $\Delta^2\lambda_\nu\geqq 0$ (that is $\{\Delta\lambda_\nu\}$ decreases to zero), we conclude that $\nu\Delta\lambda_\nu=o(1)$. Summation by parts gives

$$\sum_{\nu=0}^{n}\Delta\lambda_\nu=\sum_{\nu=0}^{n}(A_\nu^{(1)}-A_{\nu-1}^{(1)})\Delta\lambda_\nu=\sum_{\nu=0}^{n-1}A_\nu^{(1)}\Delta^2\lambda_\nu+A_n^{(1)}\Delta\lambda_n,$$

which gives

$$(4.\ 5)\qquad \sum_{\nu=0}^{\infty}A_\nu^{(1)}\Delta^2\lambda_\nu=\lambda_0,$$

since $A_n^{(1)}\Delta\lambda_\nu = O(n\Delta\lambda_n) = 0(1)$. In general, as a consequence of

$$\sum_{\nu=0}^{\infty} A_\nu^{(l-1)}\Delta^l\lambda_\nu = \lambda_0, \qquad 1 \leq l < k+1, \tag{4.6}$$

together with the fact $\Delta^{l+1}\lambda_\nu \geq 0$ we have

$$1/n \sum_{\nu=0}^{n} \nu A_\nu^{(l-1)}\Delta^l\lambda_\nu \to 0,$$[3]

$$\begin{aligned} 1/n \sum_{\nu=0}^{n} \nu A_\nu^{(l-1)}\Delta^l\lambda_\nu &\geq \Delta^l\lambda_n/n \sum_{\nu=0}^{n} \nu A_\nu^{(l-1)} = \Delta^l\lambda_n/n \sum_{\nu=0}^{n} \nu(A_\nu^{(l)} - A_{\nu-1}^{(l)}) \\ &= \Delta^l\lambda_n/n(-\sum_{\nu=0}^{n-1} A_\nu^{(l)} + nA_n^{(l)}) = (1 - A_{n-1}^{(l+1)}/nA^{(l)})A_n^{(l)}\Delta^l\lambda_n \\ &= l/(l+1)^{-1}A_n^{(l)}\Delta^l\lambda_n, \end{aligned}$$

whence we conclude that $n^l\Delta^l\lambda_n = o(1)$. Summation by parts gives

$$\sum_{\nu=0}^{n} A_\nu^{(l-1)}\Delta^l\lambda_\nu = \sum_{\nu=0}^{n} (A_\nu^{(l)} - A_{\nu-1}^{(l)})\Delta^l\lambda_\nu = \sum_{\nu=0}^{n-1} A_\nu^{(l)}\Delta^{l+1}\lambda_\nu + A_n^{(l)}\Delta^l\lambda_n,$$

which gives

$$\sum_{\nu=0}^{\infty} A_\nu^{(l)}\Delta^{l+1}\lambda_\nu = \lambda_0, \tag{4.7}$$

so that (ii) and (iii) of (4. 3) hold in general for $l = 0, 1, 2, \cdots, k$.

Finally, as is easily seen, since the terms in the (iii) of (4. 3) are positive,

$$\sum_{\nu=0}^{n} (\nu+1)^{l+s}\Delta^{l+1}\lambda_\nu = o(n^s) \tag{4.8}$$

for every $s > 0$. Thus (iv) of (4. 3) is established.

5. With reference to (4. 2), we shall prove in this section that $W_{n\,2}(\theta) = o(1)$, $W_{n\,3}(\theta) = o(1)$ uniformly in θ, and $W_{n\,1}(\theta)$ tends uniformly to a limit. Noticing the definitions of $W_{n\,2}(\theta)$ and $W_{n\,3}(\theta)$ in (4. 2), using the fact (iv) of (4. 3), we get

$$\begin{aligned} W_{n\,2}(\theta) &= O(n^{-k} \sum_{p=1}^{k} \sum_{q=1}^{p} \sum_{\nu=0}^{n-(k+1)} (n-\nu)^{k-q}(\nu+1)^{k-p+q}\Delta^{k+1-p}\lambda_{\nu+p}) \\ &= O(\sum_{p=1}^{k} \sum_{q=1}^{p} n^{-q} \sum_{\nu=0}^{n-(k+1)} (\nu+1)^{k-p+q}\Delta^{k+1-p}\lambda_{\nu+p}) = o(1), \end{aligned} \tag{5.1}$$

$$\begin{aligned} W_{n\,3}(\theta) &= O(n^{-k} \sum_{q=1}^{k} \sum_{\nu=0}^{n-(k+1)} (n-\nu)^{k-q}(\nu+1)^{q-1}\lambda_{\nu+k+1}) \\ &= O(\sum_{q=1}^{k} n^{-q} \sum_{\nu=0}^{n-(k+1)} (\nu+1)^{q-1}\lambda_{\nu+k+1}) = o(1), \end{aligned} \tag{5.2}$$

---

[3] If $\sum a_\nu$ converges, then $1/n \sum_{\nu=0}^{n} \nu a_\nu \to 0$.

since $q > 0$. Also

$$(5.3) \quad W_{n\,1}(\theta) = 1/A_n^{(k)} \sum_{p=0}^{k} C_{k+1\,p} C_{p\,0} \sum_{\nu=0}^{n-(k+1)} A_{n-\nu}^{(k)} e^{i\nu\theta} \Delta^p \nu^k \Delta^{k-p+1} \lambda_{\nu+p}$$

$$= 1/A_n^{(k)} \sum_{p=0}^{k} C_{k+1\,p} C_{p\,0} \sum_{\nu=0}^{n} A_{n-\nu}^{(k)} e^{i\nu\theta} \Delta^p \nu^k \Delta^{k-p+1} \lambda_{\nu+p} + o(1),$$

since we have added only a finite number of terms

$$1/A_n^{(k)} \sum_{p=0}^{k} C_{k+1\,p} C_{p\,0} \sum_{\nu=n-(k+1)}^{n} A_{n-\nu}^{(k)} e^{i\nu\theta} \Delta^p \nu^k \Delta^{k-p+1} \lambda_{\nu+p},$$

the order of each term is of the form $O(\nu^{k-p} \Delta^{k-p+1} \lambda_{\nu+p})$, $n - (k+1) \leq \nu \leq n$, $p = 0, 1, \cdots, k$, which is $o(1)$ for $p = 0$ (using (iii) of (4.3) with $l = k$, since $\sum_{\nu=0}^{\infty} (\nu+1)^k \Delta^{k+1} \lambda_\nu < \infty$, so that $\nu^k \Delta^{k+1} \lambda_\nu = o(\mathrm{v})$), and is $o(1/n)$ for $p = 1, \cdots, k$ (by (ii) of (4.3)).

We see that (5.3) consists of $k+1$ terms of the form

$$1/A_n^{(k)} \sum_{\nu=0}^{n} A_{n-\nu}^{(k)} e^{i\nu\theta} \Delta^p \nu^k \Delta^{k-p+1} \lambda_{\nu+p}, \qquad k \geq p,$$

which are the k-th Cesàro means of the absolutely and uniformly convergent series

$$\sum_{\nu=0}^{n} e^{i\nu\theta} \Delta^p \nu^k \Delta^{k-p+1} \lambda_{\nu+p},$$

since $\sum_{j=0}^{\infty} (\nu+1)^{k-p} \Delta^{k-p+1} \lambda_{\nu+p} < +\infty$, by virtue of (iii) of (4.3). Therefore (5.3) is uniformly convergent. (5.1), (5.2) and (5.3) together with (4.1) imply (4.2). Our theorem follows from (2.1), (3.4), (3.5) and (4.1).

The conditions $\lambda_\nu \to 0$ and $\Delta\lambda_\nu \geq 0$ do not imply the uniform $(C, 1)$ summability of the first derived series of (1.1) in the interval $(\epsilon, \pi - \epsilon)$.[4] In the same way we can show that the conditions $\lambda_\nu \to 0$ and $\Delta^k \lambda_\nu \geq 0$ do not imply the uniform (C, k) summability of the k-th derived series of (1.1).

I wish to express my gratitude to Professor Zygmund for his suggestions and encouragement.

University of California.

---

[4] A. Zygmund, *op. cit.*, p. 129, Ex. 6.

ON ISOLATED EIGENFUNCTIONS ASSOCIATED WITH BOUNDED POTENTIALS.*

By C. R. Putnam.

---

1. Let $f(t)$ be a real-valued, continuous function on the half-line $0 \leqq t < \infty$ and let λ denote a real parameter, $-\infty < \lambda < \infty$. Only real-valued solutions $x = x(t) \not\equiv 0$ of the differential equation

$$x'' + (\lambda + f(t))x = 0 \tag{1}$$

will be considered. If, for some λ, the equation (1) possesses at least one solution $x = x(t)$ not of class (L^2), that is, a solution which *fails* to satisfy

$$\int_0^\infty x^2(t)\,dt < \infty, \tag{2}$$

then, for *every* λ, the equation possesses at least one such solution; [7], p. 238. In this case, (1) is said to be in the *Grenzpunktfall* and the equation (1) and a homogeneous boundary condition

$$x(0)\cos\alpha + x'(0)\sin\alpha = 0, \qquad 0 \leqq \alpha < \pi, \tag{3}$$

determine a boundary value problem for every fixed α. By $S = S(\alpha)$ will be meant the (closed) set of λ-values constituting the spectrum of such a boundary value problem. The derivative of $S(\alpha)$, that is, the set of cluster points of $S(\alpha)$ is independent of α ([7], p. 251), and will be denoted by S'.

It is known ([7], p. 238) that if $f(t)$ is bounded, that is, if

$$|f(t)| < \text{const.}, \qquad 0 \leqq t < \infty, \tag{4}$$

or, more generally, if $f(t)$ is subject only to the unilateral restriction

$$-\infty < f(t) < \text{const.}, \qquad 0 \leqq t < \infty, \tag{5}$$

then (1) is in the Grenzpunktfall. If $f(t)$ satisfies the limit relation

$$f(t) \to 0, \qquad t \to \infty, \tag{6}$$

then the set S' is the half-line $\lambda \geqq 0$; [2], p. 71. In fact, if f is subject only to (4), it follows from [3], p. 850, that every value λ in S' satisfies the inequality $\lambda \geqq -\limsup f(t)$, where $t \to \infty$; furthermore, every λ-interval

---

* Received April 4, 1949.

$\mu_1 \leqq \lambda \leqq \mu_2$ of length not less than $\limsup f(t) - \liminf f(t)$ and for which μ_1 satisfies $\mu_1 \geqq -\limsup f(t)$, contains at least one point of S' ([5], p. 613).

If (4) is satisfied, it follows from the theorem in [9], p. 6 (cf. also [1]), that if $x(t)$ is a solution of (1) belonging to class (L^2), then $x'(t)$ also is of class (L^2). In **2**, **5**, and **6**, the following criterion for points of $S(\alpha)$ and S', in terms of the solutions of the differential equation (1), will be proved:

THEOREM (I). *Let* $f(t)$ *be a continuous function on the half-line* $0 \leqq t < \infty$ *satisfying* (4); *let* λ *denote a fixed number for which either of the inequalities*

$$(7) \qquad \lambda + \liminf_{t \to \infty} f(t) > 0; \qquad\qquad (7\text{ bis}) \quad \lambda + \limsup_{t \to \infty} f(t) < 0$$

is satisfied; finally, let $x = x(t) \not\equiv 0$ *denote any solution of* (1) *satisfying* (3) *for a fixed* α. (i) *If*

$$(8) \qquad \limsup_{t \to \infty} \int_0^t (x^2(s) + x'^2(s))\,ds/(x^2(t) + x'^2(t)) = \infty,$$

then λ *is in the set* $S(\alpha)$. (ii) *If* $x(t)$ *is of class* (L^2) *and if*

$$(9) \qquad \limsup_{t \to \infty} \int_t^\infty (x^2(s) + x'^2(s))\,ds/(x^2(t) + x'^2(t)) = \infty,$$

then λ *is in* S'.

In the proof of Theorem (I), it will be convenient to replace assumption (7) by the (apparently more restrictive but, actually, equivalent) assumption

$$(10) \qquad \lambda > \limsup_{t \to \infty} |f(t)|.$$

That (7) and (10) are equivalent follows from the fact that the differential equation (1) remains unchanged if λ and $f(t)$ are replaced by $\lambda + c$ and $f(t) - c$ ($c = \text{const.}$) respectively. Hence (cf. (4)), there is no loss of generality in supposing that

$$-\liminf_{t \to \infty} f(t) = \limsup_{t \to \infty} f(t) \qquad (= \limsup_{t \to \infty} |f(t)|);$$

and, consequently, (7) becomes identical with (10).

If (1) is in the Grenzpunktfall and if λ is not in the set S', then there exists one and (except for constant multiples) only one solution $x = y(t)$ of (1) belonging to class (L^2); [4].

As a partial corollary of Theorem (I), there will be proved

THEOREM (II). *Let* $f(t)$ *be a continuous function on the half-line* $0 \leqq t < \infty$ *satisfying* (4); *let* λ *denote a fixed number not in the set* S' *and satisfying* (7) *or* (7 bis); *finally, let* $x = y(t)$ *be a solution of* (1) *belonging to class* (L^2). *Then there exist two positive constants,* v *and* k, *satisfying*

$$y^2(t) + y'^2(t) < ve^{-kt}, \qquad 0 \leqq t < \infty. \tag{11}$$

Furthermore, if $x = x(t)$ *denotes any solution of* (1) *which is not a constant multiple of* $y(t)$, *then*

$$x^2(t) + x'^2(t) > we^{kt}, \qquad 0 \leqq t < \infty, \tag{12}$$

where w *denotes a positive constant.*

If $p(t)$ denotes a continuous periodic function on $0 \leqq t < \infty$, it is known ([10] and [6], p. 844) that the set S' associated with the differential equation

$$x'' + (\lambda + p(t))x = 0 \tag{13}$$

is identical with the region of stability of the same equation. In case $f(t)$ satisfies (4) while λ satisfies (7) or (7 bis) and is not in S', it is seen from Theorem (II) that the solutions of (1) behave as the solutions of (13) in the regions of instability; that is, in both cases, there exist (exponentially) "large" and "small" solutions.

It is known ([11], p. 604) that if $f(t)$ satisfies (5) and if λ is not in S', then the "isolated" eigenfunction $y(t)$ belonging to λ (cf. the remark preceding the statement of Theorem (II)) satisfies $y(t) = O(t^{-n})$, $t \rightarrow \infty$, for every positive constant n. According to Theorem (II), the last relation can be sharpened to the exponential estimate of (11) provided assumption (5) is strengthened to (4) and the additional assumption, either (7) or (7 bis), is made. It remains undecided, however, whether these altered hypotheses are necessary for this improved estimate.

If (4) holds and λ is arbitrary, it is known ([8], p. 391) that *any* solution $x(t)$ of (1) satisfies, for large t, the inequalities

$$e^{-k_1 t} < x^2(t) + x'^2(t) < e^{k_2 t},$$

for some pair of positive constants k_1 and k_2. It follows from a remark of Wintner [11], p. 604, that the more precise formulation of the above inequalities, given in [8], p. 391, together with (11), implies the theorem of [2] concerning S' in the case (6).

2. *Proof of Theorem* (I) *under assumption* (7 bis). Let $h(t)$

$= -(\lambda + f(t))$ and choose, in virtue of (4) and (7 bis), a constant b such that

$$(14) \qquad h(t) > b > 0, \text{ when } t \text{ is sufficiently large.}$$

Relation (7 bis) implies that (1) is non-oscillatory and consequently λ is not in S'; cf., e. g., [3]. If, therefore, $x = y(t)$ is the (essentially unique) solution of (1) belonging to class (L^2), it follows from [9] (or directly, cf. also [1]), that

$$(15) \qquad y(t) \to 0, \ y'(t) \to 0, \qquad t \to \infty.$$

The identity $(yy')' = y'^2 + yy''$ and (1) yield $(yy')' = y'^2 + hy^2$; hence, by (14), if t is sufficiently large, $y'^2 + by^2 < (yy')'$. An integration of this inequality and an application of (15) now imply

$$\int_t^\infty (y'^2 + by^2)\,ds < -y(t)y'(t),$$

if t is sufficiently large. This inequality and the inequality $|yy'| \leqq \frac{1}{2}(y^2 + y'^2)$ clearly imply

$$\limsup_{t\to\infty} \int_t^\infty (y^2 + y'^2)\,ds/(y^2(t) + y'^2(t)) < \infty.$$

This verifies the fact that the assumption (9) of (ii) is never satisfied in the case (7 bis).

On the other hand, it follows from (15) that (8) holds if $x = y(t)$ is of class (L^2). Let $x = x(t)$ be any solution of (1) linearly independent of $y(t)$. The Wronskian $x'y - xy'$ is a non-vanishing constant; consequently,

$$(16) \qquad 0 < \text{const.} = |x'y - xy'| \leqq (x^2 + x'^2)(y^2 + y'^2).$$

Hence, by (15), it is seen that $x^2 + x'^2 \to \infty$ as $t \to \infty$. As above, it is easily verified that, for large t-values,

$$\text{const.} + \int_0^t (x'^2 + bx^2)\,ds < |x(t)x'(t)|;$$

where the "const." represents a contribution of two sources, namely one related to the lower limit of integration, the other related to the fact that the inequality in (14) is assumed to be valid for large t-values. Since $|xx'| \leqq \frac{1}{2}(x^2 + x'^2) \to \infty$ as $t \to \infty$, the limit relation (8) is violated. Consequently, (8) holds if and only if $x = x(t)$ is of class (L^2). The assertion (i) is contained in this statement; and so, Theorem (I) is proved in the case (7 bis).

3. Before beginning the proofs of the remainder of Theorem (I) and of Theorem (II), it will be convenient to obtain an inequality from the "Parseval" identity for the boundary value problem determined by (1) and (3). Let $\phi(t,\lambda)=\phi(t,\lambda,\alpha)$ denote the solution of the differential equation (1) satisfying the boundary condition (3) and normalized by

$$\phi(0,\lambda)=\sin\alpha,\qquad \phi'(0,\lambda)=-\cos\alpha,$$

where the prime denotes partial differentiation with respect to t. If $\lambda=\lambda_j$ is an eigenvalue, the symbol $\phi_j(t)=\text{const.}\,\phi(t,\lambda_j)$ will denote an eigenfunction of λ_j normalized so that the integral of $\phi_j{}^2(s)$ over $0\leqq s<\infty$ is 1. If $\rho=\rho(\lambda)$ denotes the unique continuous function normalized by $\rho(0)=0$, determining the continuous spectrum of the boundary value problem (1) and (3), the eigendifferentials $dP(t,\lambda)=dP(t,\lambda,\alpha)$ are defined by

$$P(t,\lambda)=\int_0^\lambda \phi(t,\mu)\,d\rho(\mu),\ \text{i. e.,}\ dP(t,\lambda)=\phi(t,\lambda)\,d\rho(\lambda).$$

An eigenfunction ϕ_j satisfies the differential equation

$$\phi''_j+(\lambda_j+f(t))\phi_j=0, \tag{17}$$

while the eigendifferentials $dP(t,\lambda)$ satisfy $(dP)''+(\lambda+f(t))dP=0$, that is,

$$(\Delta P)''+f(t)\Delta P+\int_\lambda^{\lambda+\Delta\lambda}\mu\,d_\mu P(t,\mu)=0. \tag{18}$$

If $x(t)$ denotes any function of class (L^2) on $0\leqq t<\infty$, the Fourier "coefficients" c_j and $\Delta\Gamma(\lambda)$ are defined by

$$c_j=\int_0^\infty x(s)\phi_j(s)\,ds,\qquad \Delta\Gamma(\lambda)=\int_0^\infty x(s)\Delta P(s,\lambda,\alpha)\,ds. \tag{19}$$

The set of eigenfunctions and eigendifferentials forms a complete orthonormal system on $0\leqq t<\infty$; thus, the Parseval relation

$$\int_0^\infty x^2(s)\,ds=\sum_j c_j{}^2+\int_{-\infty}^\infty (d\Gamma)^2/d\rho \tag{20}$$

is valid.

Let $f(t)$ and $F(t)$ denote continuous functions on $0\leqq t<\infty$ satisfying (4) and

$$|F(t)|<\text{const.},\qquad 0\leqq t<\infty, \tag{21}$$

respectively. Let $g(t)$ be a continuous function of class (L^2) on $0\leqq t<\infty$. Finally, let $x(t)$ denote a solution of the equation

$$x'' + (\lambda + F(t))x = g(t) \tag{22}$$

belonging to class (L^2) and satisfying (3) for a fixed α, provided that such a solution exists. Consider the boundary value problem determined by (1) and (3) for this value α. On multiplying (22) by ϕ_j and (17) by x, subtracting the resulting equations and then integrating, it is seen that

$$(x'\phi_j - x\phi'_j)\Big|_0^t + \int_0^t (\lambda - \lambda_j + F - f)x\phi_j ds = \int_0^t g\phi_j ds.$$

Since $x(t)$ and $g(t)$ are of class (L^2) and (4), (21), hold, it is clear from (22) that the function $x'' + f(t)x$ is of class (L^2). It follows from a remark of Weyl ([7], pp. 241-242; cf. also [9] and [1]) that

$$x(t)\phi_j'(t) - x'(t)\phi_j(t) \to 0, \qquad t \to \infty.$$

Since $x(t)$ satisfies (3) it follows from (19) and the last two formula lines that

$$\int_0^\infty [(f - F)x + g]\phi_j ds = (\lambda - \lambda_j)c_j.$$

A similar calculation in which (17) is replaced by (18) shows that

$$\int_0^\infty [(f - F)x + g]\Delta P\, ds = \int_\lambda^{\lambda+\Delta\lambda} (\lambda - \mu)\, d\Gamma(\mu),$$

where $\Delta\Gamma$ is defined by (19) (cf. [5], p. 616). The last two formula lines and the Parseval relation (20) applied to the function $(f - F)x + g$ then yield

$$\int_0^\infty [(f - F)x + g]^2 ds = \sum_j (\lambda - \lambda_j)^2 c_j^2 + \int_{-\infty}^\infty (\lambda - \mu)^2 (d\Gamma)^2 d\rho(\mu).$$

In virtue of (20) and the inequality $(a + b)^2 \leqq 2(a^2 + b^2)$, it follows from the last relation that

$$2\left(\int_0^\infty (F - f)^2 x^2 ds + \int_0^\infty g^2 ds\right) \geqq m^2 \int_0^\infty x^2 ds, \tag{23}$$

where $m = m(\lambda, \alpha)$ is defined by

$$m = \min |\lambda - \mu|, \qquad \mu \text{ in } S(\alpha). \tag{24}$$

4. It will be shown that

(*) *If $T > 0$, there exist three positive constants d, c_1 and c_2 (all independent of T); a continuous function $g(t)$ on $0 \leqq t < \infty$ satisfying*

$$g(t) = 0 \text{ for } 0 \leqq t \leqq T \text{ and } T + d \leqq t < \infty; \tag{25}$$

and a function $\psi(t)$ *satisfying the differential equation*

$$\psi'' + \lambda\psi = g(t), \tag{26}$$

the relations

$$\psi(T) \neq 0, \qquad \psi'(T) = 0, \qquad \psi(T+d) = \psi'(T+d) = 0 \tag{27}$$

and, finally,

$$\int_T^{T+d} \psi^2(s)\,ds = c_1\psi^2(T), \qquad \int_0^\infty g^2(t)\,dt = c_2\psi^2(T). \tag{28}$$

Let T_1, T_2 denote a pair of numbers satisfying

$$\pi\lambda^{-\frac{1}{2}} < T_1 < T_2, \qquad T_2 - T_1 < \tfrac{1}{2}\pi\lambda^{-\frac{1}{2}} \tag{29}$$

and let $G(t)$ denote any continuous function on $0 \leqq t < \infty$ satisfying

(30) $G(t)$ is or is not 0 according as $T_1 < t < T_2$ is not or is satisfied.

If the function $\Psi(t)$ is defined by

$$\Psi(t) = \lambda^{-\frac{1}{2}} \int_{T_2}^{t} G(s) \sin \lambda^{\frac{1}{2}}(t-s)\,ds, \qquad 0 \leqq t < \infty,$$

it is seen that

$$\Psi'(t) = \int_{T_2}^{t} G(s) \cos \lambda^{\frac{1}{2}}(t-s)\,ds, \tag{31}$$

and, consequently,

$$\Psi(T_2) = \Psi'(T_2) = 0. \tag{32}$$

It is easily verified that $\Psi(t)$ satisfies the differential equation

$$\Psi'' + \lambda\Psi = G(t). \tag{33}$$

In virtue of (29) and (30) the integrand of (31) does not vanish for $T_1 < t < T_2$ and hence $\Psi'(t) \neq 0$ when $T_1 \leqq t < T_2$. It follows from (30) that on the domain $0 \leqq t \leqq T_1$ the solution $\Psi(t)$ of (33) is a non-trivial linear combination of $\sin \lambda^{\frac{1}{2}}t$ and $\cos \lambda^{\frac{1}{2}}t$. Consequently, there exists a (unique) point T_3 such that

$$\Psi(T_3) \neq 0, \qquad \Psi'(T_3) = 0, \qquad T_1 - \pi\lambda^{-\frac{1}{2}} < T_3 < T_1. \tag{32 bis}$$

Define the positive constants c_1 and c_2 by

$$c_1 = \int_{T_3}^{T_2} \Psi^2(s)\,ds/\Psi^2(T_3) \quad \text{and} \quad c_2 = \int_0^\infty G^2(s)\,ds/\Psi^2(T_3), \tag{34}$$

and d by $T_2 - T_3$.

If $T > 0$ is arbitrary, define the functions $g(t)$ and $\psi(t)$ on $0 \leqq t < \infty$ by

$$g(t) = 0 \text{ for } 0 \leqq t \leqq T \text{ and } g(t) = G(t - T + T_3) \text{ for } T \leqq t < \infty,$$

and

$$\psi(t) = \lambda^{-\frac{1}{2}} \int_{T+d}^{t} g(s) \sin \lambda^{\frac{1}{2}}(t-s)\, ds, \qquad 0 \leqq t < \infty;$$

so that T now plays the rôle of T_3 in (34) and the formula preceding it. Relations (25), (26), (27) and (28) follow from (30), (33), (32), (32 bis) and (34), in that order, and the proof of (*) is complete.

5. *Proof of* (i) *of Theorem* (I) *under assumption* (7). According to the remark following the statement of Theorem (I), it may be assumed that (10) holds. It will be shown that there exists a sequence $t_1 < t_2 < \cdots$, where $t_n \rightarrow \infty$ as $n \rightarrow \infty$, such that

$$(35) \qquad x'(t_n) = 0, \qquad n = 1, 2, \cdots,$$

and

$$(36) \qquad \int_0^{t_n} x^2(s)\, ds / x^2(t_n) \rightarrow \infty, \qquad n \rightarrow \infty.$$

In virtue of (10), a pair of constants β and S can be chosen so that

$$(37) \qquad |f(t)| < \beta < \lambda, \qquad t \geqq S.$$

Since $\lambda + f(t) > 0$, when $t \geqq S$, the graph of $x = |x(t)|$ on this domain consists of a sequence of convex arches. If $\tau_1 < \tau_2$, where $\tau_1 \geqq S$, denote two successive zeros of $x'(t)$ it is clear from (37) that

$$(38) \qquad \tau_2 - \tau_1 \leqq 2\pi(\lambda - \beta)^{-\frac{1}{2}}.$$

If $X(t) = x^2(t) + x'^2(t)$, the proof of the inequality of [8], p. 391, together with (38) and (4), shows that

$$|\log X(u_1)/X(u_2)| \leqq \gamma, \qquad u_1 \text{ and } u_2 \text{ arbitrary in } [\tau_1, \tau_2],$$

where γ is a constant (depending on λ) independent of the choice of $\tau_1 (\geqq S)$. In particular, the last formula line implies

$$x^2(\tau_2) \leqq e^{\gamma}(x^2(t) + x'^2(t)), \qquad \tau_1 \leqq t \leqq \tau_2,$$

and therefore,

$$\int_0^{t} X(s)\, ds / X(t) \leqq e^{\gamma} \int_0^{\tau_2} X(s)\, ds / x^2(\tau_2), \qquad \tau_1 \leqq t \leqq \tau_2.$$

The last inequality and (8) clearly imply the existence of a sequence $t_1 < t_2 < \cdots$, where $t_n \to \infty$, as $n \to \infty$, such that (35) and

$$\int_0^{t_n} (x^2 + x'^2)\, ds / x^2(t_n) \to \infty, \qquad n \to \infty,$$

hold. Multiplication of (1) by x followed by an integration and an application of (35) shows that for $n = 1, 2, \cdots$,

$$\int_0^{t_n} x'^2(s)\, ds = -x(0)x'(0) + \int_0^{t_n} (\lambda + f) x^2 ds. \tag{39}$$

In virtue of (4), the last two relations imply $(A + B \int_0^{t_n} x^2(s) ds)/x^2(t_n) \to \infty$, $n \to \infty$, where A and B are positive constants. This last relation obviously implies (36).

Let D denote an arbitrary positive constant and choose a number N, depending on D, such that (cf. (37))

$$t_{N-1} > S \tag{40}$$

and (cf. (36))

$$\int_0^{t_n} x^2(s)\, ds / x^2(t_N) > D. \tag{41}$$

Let s_N denote the first zero of $x(t)$ to the left of t_N and choose u_N, where $s_N < u_N < t_N$, so near t_N that

$$| x'(u_N) | \leqq | x(u_N) | (\lambda - \beta)^{\frac{1}{2}} / \pi, \tag{42}$$

where β is defined by (37), and

$$\int_0^{u_N} x^2(s)\, ds / x^2(u_N) > D/2. \tag{43}$$

Let $\xi(t)$ denote the solution of the differential equation

$$\xi'' + (\lambda + \beta)\xi = 0 \tag{44}$$

which satisfies

$$\xi(u_N) = x(u_N), \qquad \xi'(u_N) = x'(u_N). \tag{45}$$

Let R $(> u_N)$ denote the first zero of the function $\xi'(t)$ to the right of u_N and define a continuous function $F = F(t)$ on $0 \leqq t < \infty$ so as to satisfy

$$F(t) = f(t) \text{ for } 0 \leqq t \leqq u_N, \qquad | F(t) | \leqq | f(t) | \text{ for } u_N \leqq t \leqq R, \\ F(t) = 0 \text{ for } R \leqq t < \infty. \tag{46}$$

Let $y = y(t)$ denote the solution of the differential equation

$$y'' + (\lambda + F(t))y = 0 \tag{47}$$

which satisfies

$$y(u_N) = x(u_N), \qquad y'(u_N) = x'(u_N). \tag{48}$$

In virtue of (37), (40) and (46) it follows that

$$\lambda - \beta < \lambda + F(t) < \lambda + \beta, \qquad u_N \leqq t < \infty. \tag{49}$$

Hence, if T denotes the first zero of the function $y'(t)$ to the right of u_N, relations (44) to (49) imply the inequalities

$$u_N < R < T. \tag{50}$$

It is easily verified, as a consequence of (48) and (49) that

$$|\, y(T) \,| \leqq |\, x(u_N) \,| + |\, x'(u_N) \,| \, (T - u_N) \text{ and } T - u_N \leqq \pi(\lambda - \beta)^{-\frac{1}{2}}.$$

The last two relations and (42) imply $|\, y(T) \,| \leqq 2 \,|\, x(u_N) \,|$. Since (46), (47) and (48) show that $y(t) \equiv x(t)$ for $0 \leqq t \leqq u_N$, it follows from (43) and (50) that

$$\int_0^T y^2(s)\, ds / y^2(T) > D/8. \tag{51}$$

Identify the point T, just constructed, with that occuring in the italicized statement (*) of this section. Since the assertions of (*) remain unchanged if $g(t)$ and $\psi(t)$ are replaced by $Cg(t)$ and $C\psi(t)$ respectively, where C is any non-vanishing constant, it may be supposed that the function $\psi(t)$ satisfies

$$\psi(T) = y(T). \tag{52}$$

In virtue of (27) and the definition of T (in the last paragraph) it follows that

$$y'(T) = \psi'(T) = 0.$$

Define the function $z = z(t)$ on $0 \leqq t < \infty$ by

$$z(t) = y(t) \text{ for } 0 \leqq t \leqq T \text{ and } z(t) = \psi(t) \text{ for } T \leqq t < \infty.$$

The statement (*) together with (46), (49) and the properties of $y(t)$ show that $z(t)$ satisfies the differential equation

$$z'' + (\lambda + F(t))z = g(t) \tag{53}$$

and boundary condition (3), for the (fixed) α specified in the statement of Theorem (I). The third relation of (27) and the facts that $g(t) = 0$ when $T + d \leqq t < \infty$ and $F(t) = 0$ when $t \geqq R$, where $R < T < T + d$, imply that

$$(54) \qquad z(t) \equiv 0, \qquad T + d \leqq t < \infty.$$

Let the functions $f(t)$, $F(t)$ and $g(t)$, considered above, be identified with those appearing in the latter part of **3**. It follows from (23), (46), (53) and (54) that

$$(55) \qquad 2\left(\int_{u_N}^{T+d} (F-f)^2 z^2 ds + \int_0^\infty g^2 ds\right) \geqq m^2 \int_0^\infty z^2 ds.$$

Hence, by (52) and the second equality of (28), $\int_0^\infty g^2 ds = c_2 y^2(T)$, and by (4) and (46),

$$(F(t) - f(t))^2 < 4f^2(t) < c_3 \ (= \text{const.}), \qquad 0 \leqq t < \infty.$$

Thus $\int_T^{T+d} (F-f)^2 z^2 dt \leqq c_3 c_1 y^2(T)$, by the first equality of (28) and the definition of $z(t)$. Since $T - u_N \leqq \pi(\lambda - \beta)^{-\frac{1}{2}}$ $(= c_4)$ and $|z(t)| = |y(t)| \leqq |y(T)|$ for $u_N \leqq t \leqq T$, by the definition of u_N and T, it follows that

$$\int_{u_N}^{T} (F-f)^2 z \, ds \leqq c_3 c_4 y^2(T).$$

Relation (55) and the above relations imply

$$c_5 y^2(T) \geqq m^2 \int_0^\infty z^2 ds \geqq m^2 \int_0^T y^2 ds,$$

where c_5 denotes the constant $c_5 = 2(c_2 + c_1 c_3 + c_3 c_4)$. It follows therefore from (51) that $m^2 D$ is less than a constant which is independent of D. Consequently, $m = 0$, since D may be chosen arbitrarily large. This means, according to the definition (24) of m, that λ is in $S(\alpha)$ and the proof of (i) is complete.

6. *Proof of* (ii) *of Theorem* (I) *under assumption* (7). By a process similar to that used in **5**, it is easily shown that there exists a sequence $t_1 < t_2 < \cdots$, where $t_n \rightarrow \infty$ as $n \rightarrow \infty$, such that (35) and

$$\int_{t_n}^\infty (x^2 + x'^2) ds / x^2(t_n) \rightarrow \infty, \qquad n \rightarrow \infty.$$

In virtue of (4), relations (15) are valid for $x=y$ and consequently, from (35),

$$\int_{t_n}^{\infty} x'^2 ds = \int_{t_n}^{\infty} (\lambda + f) x^2 ds, \qquad n=1,2,\cdots,$$

cf. (39). The last two formula lines clearly imply $\int_{t_n}^{\infty} x^2 ds / x^2(t_n) \to \infty$, $n \to \infty$, corresponding to (36). The remainder of the proof of (ii), including the construction of functions corresponding to $g(t)$, $F(t)$, etc. occurring in **4** and **5** is similar to that of (i) provided only that it is observed that the rôles played by 0 and ∞ there may be interchanged in the present case. This procedure is permitted by the assumption that x (and hence x') is of class (L^2), together with the resulting implication (15) for $x=y$. A copying of the proof of (i) given in the last two sections, with the appropriate modifications as indicated, leads to the construction of a function $Z(t)$, corresponding to the function $z(t)$ above, satisfying the identity $Z(t) \equiv 0$, $0 \leqq t \leqq \text{const.} < \infty$, corresponding to (54) for $z(t)$. Since $Z(t)$ satisfies, therefore, the boundary condition (3) for *every* α, the function $m = m(\lambda, \alpha)$ defined by (24) satisfies the identity $m(\lambda, \alpha) \equiv 0$, $0 \leqq \alpha < \pi$. That is, λ is in $S(\alpha)$ for every α and hence, in S'. This completes the proof of (ii) (and hence of Theorem (I)).

7. *Proof of Theorem* (II). It follows from (ii) of Theorem (I) that the expression appearing on the left of the equality (9), in which $x=y$, is finite. Hence, if $Y(t)$ is defined by

$$(56) \qquad Y(t) = \int_t^{\infty} (y^2 + y'^2)\, ds,$$

there exists a positive constant k such that $Y(t)/-Y'(t) < 1/k$, $0 \leqq t < \infty$; consequently,

$$(57) \qquad Y(t) < Y(0) e^{-kt}.$$

Relations (15) and (1), where $x = y(t)$, imply that

$$y'^2(t) + \lambda y^2(t) = 2 \int_t^{\infty} (y'y'' + \lambda y'y)\, ds = -2 \int_t^{\infty} f y y' ds.$$

In virtue of (4), the last relation and the inequality $|yy'| \leqq 2(y^2 + y'^2)$ imply

$$y'^2(t) + \lambda y^2(t) \leqq \text{const.} \int_t^{\infty} (y^2 + y'^2)\, ds, \qquad 0 \leqq t < \infty.$$

The relation (11) follows from this inequality, (56), (57) and $\lambda > 0$.

Relation (12) clearly follows from (11) and (16), and the proof of Theorem (II) is complete. (The existence of a constant $k > 0$ satisfying (12) can also be obtained directly from (i) of Theorem (I).)

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] P. Hartman, "The L^2-solutions of linear differential equations of second order," *Duke Mathematical Journal*, vol. 14 (1947), pp. 323-326.

[2] ———, "On the spectra of slightly disturbed linear oscillators," *American Journal of Mathematics*, vol. 71 (1949), pp. 71-79.

[3] ——— and C. R. Putnam, "The least cluster point of the spectrum of boundary value problems," *ibid.*, vol. 70 (1948), pp. 849-855.

[4] ——— and A. Wintner, "An oscillation theorem for continuous spectra," *Proceedings of the National Academy of Sciences*, vol. 33 (1947), pp. 376-379.

[5] C. R. Putnam, "The cluster spectra of bounded potentials," *American Journal of Mathematics*, vol. 71 (1949), pp. 612-620.

[6] S. Wallach, "The spectra of periodic potentials," *ibid.*, vol. 70 (1948), pp. 842-848.

[7] H. Weyl, "Ueber gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürlicher Funktionen," *Mathematische Annalen*, vol. 68 (1910), pp. 222-269.

[8] A. Wintner, "The adiabatic linear oscollator," *American Journal of Mathematics*, vol. 68 (1946), pp. 385-397.

[9] ———, "(L^2)-connections between the potential and kinetic energies of linear systems," *ibid.*, vol. 69 (1947), pp. 5-13.

[10] ———, "Stability and spectrum in the wave mechanics of lattices," *Physical Review*, vol. 72 (1947), pp. 81-82.

[11] ———, "On the smallness of isolated eigenfunctions," *American Journal of Mathematics*, vol. 71 (1949), pp. 603-611.

ON THE DERIVATIVES OF THE SOLUTIONS OF ONE-DIMENSIONAL WAVE EQUATIONS.*

By PHILIP HARTMAN and AUREL WINTNER.

1. On the half-line $0 \leqq t < \infty$, consider the linear differential equation

$$x'' + q(t)x = 0, \tag{1}$$

where $q(t)$ is real-valued and continuous. By solutions of (1) will be meant real-valued solutions $x(t) \not\equiv 0$. For such a solution, and its derivative, the (L^2)-character is defined by the respective conditions

$$(2_1) \quad \int_0^\infty x^2(t)\,dt < \infty; \qquad (2_2) \quad \int_0^\infty x'^2(t)\,dt < \infty.$$

The present note centers about the following facts:

For an unspecified $q(t)$ *in* (1), *where* $0 \leqq t < \infty$, *let* $x = x(t)$ *and* $x = y(t)$ *be two linearly independent solutions. Then*

(i) *for suitable* $q(t)$, *both* $x(t)$ *and* $y(t)$ *become of class* (L^2);

(ii) *for no* $q(t)$ *can* $x'(t)$ *and* $y'(t)$ *be of class* (L^2);

(iii) *if* $x'(t)$ *is of class* (L^2), *then* $y(t)$ *cannot be of class* (L^2).

Remark. If (1) is generalized to

$$(p(t)x')' + q(t)x = 0, \tag{3}$$

where $p(t)$ is positive and continuous, then (ii) cannot be asserted. For instance, if $p(t) = e^t$ and $q(t) = e^{-t}$, then every solution of (3), being given by $x(t) = c_1 \cos e^{-t} + c_2 \sin e^{-t}$, has a derivative satisfying (2_2). What is true is that, with reference to (3),

$$\int_0^\infty p(t)x'^2(t)\,dt = \infty \text{ holds for some } x(t) \text{ if } \int_0^\infty \{p(t)\}^{-1}dt = \infty. \tag{4}$$

Since (ii) refers to (1), where $p(t) \equiv 1$, it is clear that (ii) is contained

---

* Received May 21, 1949.

in (4). Conversely, (4) can be obtained from (ii) by a change of the independent variable.

Ad (i). The asymptotic results of [10] imply that, if $q(t)$ is of class C'', tends to ∞ as $t \rightarrow \infty$, and satisfies

$$\limsup_{t\rightarrow\infty} q'^2/q^3 < \infty \text{ and } \int^{\infty} (q'^2/q^{5/2})\,dt < \infty,$$

then every or no solution of (1) is of class (L^2) according as $\int^{\infty} \{q(t)\}^{-\frac{1}{2}}dt$ is convergent or divergent. On the other hand, both requirements of the last formula line are readily seen to be satisfied whenever $q(t)$ is a logarithmico-exponential function which tends to ∞ as $t \rightarrow \infty$. Accordingly, if $q(t)$ is any such function, (i) will hold whenever $q(t)$ increases at least as fast as t^2 (or just $t^2/\log^2 t, \cdots$); cf. [4], p. 306.

Proof of (ii). Since $x(t)$ and $y(t)$ are linearly independent solutions of (1), their Wronskian is a non-vanishing constant. Hence, it can be assumed that

$$xy' - yx' \equiv 1. \tag{5}$$

Then, by Schwarz's inequality, $1 \leqq (x^2 + y^2)(x'^2 + y'^2)$. Consequently, if (ii) is assumed to be false, i.e., if both $x'(t)$ and $y'(t)$ are of class (L^2), it follows that $1/(x^2 + y^2)$ is absolutely integrable (over $0 \leqq t < \infty$). This implies that (1) is non-oscillatory; cf. [5], pp. 210-211. But if (1) is non-oscillatory, then it must possess some solution $x = z(t)$ satisfying

$$\int_T^{\infty} \{z(t)\}^{-2}dt < \infty \text{ if } T \text{ is large enough;} \tag{6}$$

cf. [2], p. 703. On the other hand, since $z(t)$ is a linear combination of $x(t)$ and $y(t)$, and since $x'(t)$ and $y'(t)$ are supposed to be of class (L^2), the function $z'(t)$ is of class (L^2). In view of (6) and of Schwarz's inequality, this implies that the product $z'z^{-1}$ is integrable over the half-line $T \leqq t < \infty$. Since $z'z^{-1} = (\log z)'$, it follows that $\log z(t)$, and therefore $z(t)$ itself, tends to a finite limit as $t \rightarrow \infty$. Since this contradicts (6), the proof of (ii) is complete.

Proof of (iii). Suppose that (iii) is false, i.e., that $x'(t)$ and $y(t)$ are of class (L^2). Then, if (5) is written in the form

$$(xy)' - 1 = 2x'y, \tag{7}$$

it is seen that $(xy)' - 1$ is (absolutely) integrable over $0 \leqq t < \infty$. Hence, $xy - t$ tends to a finite limit as $t \rightarrow \infty$. This implies that, if t is large enough, $x(t)$ and $y(t)$ do not vanish and satisfy $1/|x(t)| < |y(t)|$. Since $y(t)$ is of class (L^2), it follows that $1/x^2(t)$ is absolutely integrable over $T \leqq t < \infty$, if T is large enough (so large that $x(t) \neq 0$ when $t \geqq T$). But (5) shows that $1/x^2(t)$ is identical with the derivative of y/x. Consequently, y/x tends to a finite limit as $t \rightarrow \infty$. This limit is 0; for, on the one hand, y is of class (L^2) and, on the other hand, $xy \sim t$ as $t \rightarrow \infty$. Hence

$$y(t)/x(t) = -\int_t^\infty ds/x^2(s) \text{ whenever } t \geqq T.$$

Thus, $yx < 0$ if $t \geqq T$. But this contradicts $xy \sim t$, and completes therefore the proof of (iii).

2. A modification of this proof leads to the substantial refinement of (iii). In this connection, the following lemma is of interest:

LEMMA. *If (1), where $0 \leqq t < \infty$, is non-oscillatory, and if $x = x(t)$ and $y = y(t)$ are two linearly independent solutions the second of which is of class (L^2), then (6) is satisfied by $z = x$.*

This Lemma is between the lines of [2], p. 703. It can be verified as follows: Since (1) is non-oscillatory, $x(t) \neq 0$ if t is large enough, say $t \geqq T$. Then $(y/x)' = 1/x^2$, by (5). Hence, if (6) is denied for $z = x$, it follows that $y/x \rightarrow \infty$. But this is impossible, since y is of class (L^2) while x is not. For, if x were of class (L^2), then, since y is, (1) had two linearly independent solutions of class (L^2). As shown in [2], this is contradicted by the assumption that (1) is non-oscillatory.

The refinement of (iii), referred to above, is as follows:

(*) *If (1), where $0 \leqq t < \infty$, has a solution $x(t)$ the derivative of which satisfies*

$$\int_0^t x'^2(s)\,ds = O(t^2), \tag{8}$$

then no solution linearly independent of this $x(t)$ is of class (L^2) (while $x(t)$ itself may, but need not, be of class (L^2)).

COROLLARY. *If the derivative of some solution $x(t) \not\equiv 0$ of (1) satisfies (8) (e.g., if*

$$\text{(8 bis)} \qquad \text{either } x'(t) = O(t^{\frac{1}{2}}) \text{ or } \int_0^\infty x'^2(t)\,dt < \infty$$

holds for some $x(t) \not\equiv 0$), *then* (1) *cannot have two linearly independent solutions of class* (L^2).

Remark. The $O(t^2)$ in (8) cannot be improved to $O(t^{2+\epsilon})$, where $\epsilon > 0$. In fact, the $O(t^{\frac{1}{2}})$ in (8 bis) cannot be relaxed to $O(t^{\frac{1}{2}+\epsilon})$. In order to see this, it is sufficient to choose $q(t) = t^{2+\epsilon}$ and then apply the general asymptotic results of [10].

Proof of (*). Suppose that (*) is false. Then (1) has a solution $x = y(t)$ which is of class (L^2) and satisfies (5), where $x(t)$ is the solution occurring in (8). Hence, if C and T are large enough,

$$\int_0^t x'^2(s)\,ds < (Ct)^2 \text{ if } T < t < \infty, \text{ and } \int_T^\infty y^2(s)\,ds < (4C)^{-2}.$$

Consequently, Schwarz's inequality shows that, if $T > 1$,

$$2\int_T^t |x'(s)y(s)|\,ds < \tfrac{1}{2}t \text{ whenever } t > T.$$

Since (7) implies that $x(t)y(t) = \text{const.} + 2\int_T^t x'(s)y(s)\,ds + t$, it follows that

$$\text{(9)} \qquad x(t)y(t) > \tfrac{1}{2}t + \text{const. if } T < t < \infty.$$

Since $y(t)$ is of class (L^2), it is clear from (9) that $x(t)$ cannot be of class (L^2). It also follows from (9) that $x(t)$, $y(t)$ do not vanish for large t, i. e., that (1) is non-oscillatory. Hence, by the Lemma, (6) is satisfied by $z(t) = x(t)$. Clearly, the balance of the proof of (*) is substantially identical with the end of the above Proof of (iii).

(* bis) *The assertion of* (*) *remains true if* (1) *is generalized to* (3) *but* (8) *is replaced by*

$$\text{(10)} \qquad \int_0^t x'^2(s)\,ds = O\Big(\int_0^t \{p(s)\}^{-1}ds\Big)^2,$$

provided that the integral on the right of (10) *is not* $O(1)$; cf. (4).

COROLLARY. *If* (3), *where* $0 \leqq t < \infty$, *has a solution the derivative of which satisfies* (10), *and if*

$$\int_0^\infty \{p(s)\}^{-1} ds = \infty, \tag{4 bis}$$

then (3) *cannot have two linearly independent solutions of class* (L^2).

3. Let (1) now be replaced by

$$x'' + \{q(t) + \lambda\}x = 0, \tag{11}$$

where λ is a real parameter. Suppose that (11) is of *Grenzpunkt* type, i. e., that not every solution of (11) is of class (L^2). In order that this be the case, it is sufficient that

$$-\infty \leqq \limsup q(t) < \infty \qquad (t \to \infty) \tag{12}$$

([8], p. 238). Another sufficient condition is the existence of a constant satisfying the unilateral Lipschitz condition

$$q(t_2) - q(t_1) < \text{const.}\ (t_2 - t_1) \text{ for } 0 \leqq t_1 < t_2 < \infty \tag{13}$$

([4], p. 296). The above considerations lead to a more far-reaching result.

In order to describe the situation completely, assign to (11) a linear boundary condition at $t = 0$, e. g., $x(0) = 0$ or, more generally,

$$x(0) \cos \alpha + x'(0) \sin \alpha = 0, \qquad (0 \leqq \alpha < \pi). \tag{14}$$

Denote by $S(\alpha)$ the set of λ-values representing the spectrum which is determined by (11) and (14) when (11) is of *Grenzpunkt* type. According to Weyl [8], p. 251, the set of the cluster values of $S(\alpha)$ is independent of α and can, therefore, be denoted simply by S'. Since $S(\alpha)$ is closed, S' is contained in every $S(\alpha)$. Conversely, a λ is in S' whenever it is in every $S(\alpha)$ (or, for that matter, in $S(\alpha_1)$ and $S(\alpha_2)$, where $\alpha_1 \not\equiv \alpha_2 \bmod \pi$). Let S' be called the essential spectrum of (11).

The following theorem can now be proved:

(I) *Suppose that there exists a* $\lambda = \lambda_0$ *corresponding to which* (11) *has a solution* $x(t) \not\equiv 0$ *satisfying* (8) (*e. g.*, (8 bis)). *Then* (11) *is of Grenzpunkt type. Furthermore, either* $x(t)$ *is of class* (L^2) *or* λ_0 *is in the essential spectrum.*

The second of these two possibilities (which are not mutually exclusive) must occur if (8) *is satisfied by two linearly independent solutions of the case* $\lambda = \lambda_0$ *of* (11).

Proof of (I). According to Weyl [8], p. 238, all solutions of (11) are of class (L^2) for some λ only if the same is true for every λ. Since there is no loss of generality in assuming that $\lambda_0 = 0$, the first assertion of (I) is equivalent to the Corollary of (*).

In order to prove the second assertion of (I), suppose that $\lambda_0 = 0$ is not in the essential spectrum. Then (1) must have a solution of class (L^2); cf. [3]. This solution, say $x = x^*(t)$, is unique to a constant factor, since (1) cannot have two linearly independent solutions of class (L^2). It remains to show that $x^*(t)$ is linearly dependent on the $x(t)$ for which (8) is assumed. But if such were not the case, an application of (*) to $y(t) = x^*(t)$ would lead to a contradiction.

This proves the second assertion of (I). Clearly, the third assertion of (I) follows in the same way as the second. In fact, (*) implies that no (non-trivial) solution of (11) is of class (L^2). Hence, by [3], the value $\lambda = \lambda_0$ is in the essential spectrum.

4. A modification of the preceding arguments leads to the following theorem:

(II) *Suppose that the coefficient function of* (11) *satisfies* (12) (*so that, in particular,* (12) *is of Grenzpunkt type*). *Suppose further that* (12) *has, for some* $\lambda = \lambda_0$, *a solution* $x = x(t) \not\equiv 0$ *satisfying*

$$\int_0^t x^2(s)\,ds = O(t^N) \text{ for some } N. \tag{15}$$

Then either $x(t)$ *is of class* (L^2) *or* λ_0 *is in the essential spectrum.*

The second of these two possibilities (which are not mutually exclusive) must occur if (15) *is satisfied by two linearly independent solutions of the case* $\lambda = \lambda_0$ *of* (11).

The assertions of (II) were proved in [6] in the particular case in which $x(t) = O(1)$, rather than just (15), is assumed.

Proof of (II). Without loss of generality, let $\lambda_0 = 0$. Then (1) is supposed to have a solution $x(t) \not\equiv 0$ satisfying (15). If $x = y(t)$ is any solution of (1) linearly independent of this $x(t)$, then, since the Wronskian $xy' - xy'$ is a non-vanishing constant, the square of this constant will satisfy

$$0 < \text{const.} \leqq (x^2 + x'^2)(y^2 + y'^2) \text{ for } 0 \leqq t < \infty,$$

by Schwarz's inequality. On the other hand, from (15),

$$\int_1^\infty t^{-K} x^2(t)\,dt < \infty$$

holds for some K (in fact, for every $K > N + 1$). But the existence of such a K and the assumption (12) imply that

$$\int_1^\infty t^{-K} x'^2(t)\,dt < \infty$$

holds for the same K; cf. [9], p. 9. In view of the last three formula lines, there exists a $C > 0$ satisfying

$$\int_1^\infty \{y^2(t) + y'^2(t)\}^{-1} t^{-C}\,dt < \infty. \tag{16}$$

Suppose that the first assertion of (II) is false, i. e., that $x(t)$ is not of class (L^2) and $\lambda = 0$ is not in the essential spectrum. Then (1) has a solution $x = y(t)$ which is linearly independent of $x(t)$ and is of class (L^2); cf. [3]. But (12) necessitates, for every such $y(t)$, the estimate $y(t) = O(t^{-n})$, where n is arbitrarily large; cf. [11]. It follows therefore by the procedure of [1] that $y'(t) = O(t^{-n})$ holds for every n. This pair of O-estimates contradicts (16), since n can be chosen large enough with reference to a fixed C. This contradiction proves the first assertion of (II).

In the remaining assertion of (II), the assumption is that (15) holds for two linearly independent solutions of (1), say for $x = x(t)$ and $x = y(t)$. Since (16) was deduced from (15), and since $x(t)$ in (15) can now be replaced by $y(t)$, it follows that both (16) and (16′) hold in the present case, if (16′) denotes what results if $y(t)$ in (16) is replaced by $x(t)$. Consequently, (1) has no (non-trivial) solution $x = x(t)$ satisfying the estimates $x(t) = O(t^{-n})$ and $x'(t) = O(t^{-n})$, where n is arbitrarily large. Hence, by

[11], the point $\lambda = 0$ is in the essential spectrum and the proof of (II) is complete.

Remark. If the assumption (12) of (II) is strengthened to

$$(17) \qquad |q(t)| < \text{const.} \qquad (0 \leqq t < \infty)$$

and if λ in (11) satisfies

$$(18) \qquad |\lambda| > \limsup |q(t)| \qquad (t \to \infty),$$

then (II) remains correct when (15) is replaced by

$$(19) \qquad \int_0^t x^2(s)\,ds = O(e^{\epsilon t}) \text{ for every } \epsilon > 0.$$

The proof of this remark is similar to that of (II). For, by (19),

$$\int_0^\infty x^2(t)e^{-\epsilon t}dt < \infty$$

for $\epsilon > 0$, and by the arguments of [9], p. 9,

$$\int_0^\infty x'^2(t)e^{-\epsilon t}dt < \infty$$

holds in view of (12). On the other hand, it is shown in [7] that (17) and (18) imply that if $\lambda = \lambda_0$ is not in the essential spectrum, then (11) has a solution $x = y(t)$ satisfying the estimates $y(t) = O(e^{-kt})$ and $y'(t) = O(e^{-kt})$ for some $k > 0$. The proof can now be completed as above.

It remains undecided whether or not the assertion (II) modified by replacing (15) by (19) is true without the additional assumptions (17) and (18).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] P. Hartman, "The L^2-solutions of linear differential equations of second order," *Duke Mathematical Journal*, vol. 14 (1947), pp. 323-326.

[2] ———, "Differential equations with non-oscillatory eigenfunctions," *ibid.*, vol. 15 (1948), pp. 697-709.

[3] ——— and A. Wintner, "An oscillation theorem for continuous spectrum," *Proceedings of the National Academy of Sciences*, vol. 33 (1947), pp. 376-379.

[4] ——— and A. Wintner, "Criteria of non-degeneracy of the wave equation," *American Journal of Mathematics*, vol. 70 (1948), pp. 295-308.

[5] ——— and A. Wintner, "A criterion for the non-degeneracy of the wave equation," *ibid.*, vol. 71 (1949), pp. 206-213.

[6] ——— and A. Wintner, "On the location of spectra of wave equations," *ibid.*, vol. 71 (1949), pp. 214-217.

[7] C. R. Putnam, "On isolated eigenfunctions associated with bounded potentials," *ibid.*, vol. 72 (1950), pp. 135-147.

[8] H. Weyl, "Ueber gewöhnliche Differentialgleichungen mit Singularitäten und die zugenhörigen Entwicklungen willkürlicher Funktionen," *Mathematische Annalen*, vol. 68 (1910), pp. 222-269.

[9] A. Wintner, "(L^2) connections between the potential and kinetic energies of linear systems," *American Journal of Mathematics*, vol. 69 (1947), pp. 5-13.

[10] ———, "On the normalization of characteristic differentials in continuous spectra," *Physical Review*, vol. 72 (1947), pp. 516-517.

[11] ———, "On the smallness of isolated eigenfunctions," *American Journal of Mathematics*, vol. 71 (1949), pp. 603-611.

ZUSÄTZLICHE STABILITÄTSBETRACHTUNG BETREFFEND "DIE SYMMETRISCHEN PERIODISCHEN BAHNEN DES RESTRINGIERTEN DREIKÖRPERPROBLEMS IN DER NACHBARSCHAFT EINES KRITISCHEN KEPLERKREISES." *

Von ERNST HÖLDER.

Einer Anregung von Wintner aus dem Jahr 1938 zufolge habe ich (im Sommer 1945) auch die *Stabilität* der in meiner im Hill-Gedächtnisheft erschienenen Arbeit[1] berechneten periodischen Bahnen untersucht. Dazu ist Λ^* bis zu den Gliedern *vierten Grades* in $\zeta, \zeta', \cdots$ explizit zu berechnen. Man hat jetzt also vollständiger

$$(21') \qquad V_4 = V_3 + 2a^{-1}\zeta^4 = V_3 + a^2m^{-2}\cdot 2(m+1)^2\zeta^4.$$

In den beiden ersten Zeilen von (23) erscheint jetzt die Entwicklung des Faktors

$$[1 + 2\zeta + \zeta^2 + \zeta'^2]^{\frac{1}{2}} = (1 + \zeta + \tfrac{1}{2}\zeta'^2 - \tfrac{1}{2}\zeta\zeta'^2 + \tfrac{1}{2}\zeta^2\zeta'^2 - \tfrac{1}{8}\zeta'^4 + \cdots)$$

sowie statt $V_3{}^{\frac{1}{2}}$

$$(24') \quad V_4{}^{\frac{1}{2}} = V_3{}^{\frac{1}{2}} + am^{-1}\{-\tfrac{1}{8}(m+1)(17m^3 + 15m^2 + 3m - 3)\zeta^4 + \cdots\}.$$

Die Restglieder $\Lambda^{(1)}, \Lambda^{(2)}, \cdots$ enthalten ausser von ζ, ζ' unabhängigen Gliedern solche Glieder dritter Ordnung in κ, μ, ζ, ζ', die den Faktor μ enthalten, Glieder vierter Ordnung, die den Faktor κ enthalten, ferner Glieder fünfter und höherer Ordnung.

In

$$(27) \qquad a^{-2}m^2\Lambda^* = L$$

braucht man ausser

$$\kappa\Theta^{(2)}(\zeta, \zeta) = \kappa[B_{1020}\zeta^2 + B_{1002}\zeta'^2] \text{ mit } B_{1020} = {}^3\!/_2 j(3m+1),\ B_{1002} = \tfrac{1}{2}$$

und

$$\Theta^{(3)}(\zeta, \zeta, \zeta) = B_{0030}\zeta^3 + B_{0012}\zeta\zeta'^2 \text{ mit } B_{0030} = -j^2m,$$
$$B_{0012} = -j = -(m+1)$$

---

* Received June 2, 1949.

[1] E. Hölder, "Die symmetrischen periodischen Bahnen des restringierten Dreikörperproblems in der Nachbarschaft eines kritischen Keplerkreises," *American Journal of Mathematics*, vol. 60 (1938), pp. 801-814.

noch explizit die weiteren Glieder

$$\Theta^{(4)}(\zeta,\zeta,\zeta,\zeta) = B_{0040}\zeta^4 + B_{0022}\zeta^2\zeta'^2 + B_{0004}\zeta'^2 .$$

mit

$$B_{0040} = -j^2(17m^2+6m+1)/8, \quad B_{0022} = j(-m+5)/4, \quad B_{0004} = -1/8.$$

Die Lagrange-Funktion wird dann vollständiger

$$(27') \quad L = a^{-2}m\Lambda^* = \tfrac{1}{2}\zeta'^2 - \tfrac{1}{2}j^2\zeta^2 + \Sigma B_{ijkl}\kappa^i\mu^j\zeta^k\zeta'^l = L^{(2)}(\zeta,\zeta) + \Theta$$
$$= L^{(2)}(\zeta,\zeta) + \kappa\Theta^{(1)}(\zeta) + \kappa^2\bar{\Theta}^{(1)}(\zeta) + \kappa\Theta^{(2)}(\zeta,\zeta) + \Theta^{(3)}(\zeta,\zeta,\zeta)$$
$$+ \Theta^{(4)}(\zeta,\zeta,\zeta,\zeta) + \mu L^{(1)}(\zeta) + \Lambda^{(4)}.$$

Die in ζ, ζ' linearen Glieder $\kappa\Theta^{(1)}(\zeta) + \kappa^2\bar{\Theta}^{(1)}(\zeta) + \mu L^{(1)}(\zeta)$ interessieren uns im folgenden nicht.

Die lineare Variationsgleichung für einen infinitesimalen Zuwachs

$$Z(\tau) = C_1 \cos j\tau + C_2 \sin j\tau + \cdots = C_1Z_1(\tau) + C_2Z_2(\tau) + \cdots$$

lautet

$$(L_{\zeta'\zeta'}Z' + L_{\zeta'\zeta}Z)' - (L_{\zeta\zeta'}Z' + L_{\zeta\zeta}Z) = 0.$$

Die an dieser Stelle nicht zu entwickelnde Störungstheorie der charakteristischen Exponenten (analog zu jener der Eigenwerte) liefert einen Instabilitätsbereich zwischen den beiden Kurven in der (κ, ξ)-Ebene, auf denen die Variationsgleichung eine periodische Lösung besitzt. Dafür muss die "rechte Seite" die beiden Orthogonalitätsrelationen bezüglich $Z_i(\tau)$ erfüllen, die für C_1, C_2 zwei lineare homogene Gleichungen

$$-\sum_{j=1}^{2}\int_0^{2\pi}(\Theta_{\zeta'\zeta'}Z'_iZ'_j + \Theta_{\zeta\zeta'}(Z_iZ'_j + Z'_iZ_j) + \Theta_{\zeta\zeta}Z_iZ_j)\,d\tau.\ C_j = 0. \qquad (i=1,2)$$

darstellen. Für deren Lösbarkeit ist das Verschwinden der Determinante

$$(*) \qquad \det\Big(\int_0^{2\pi}(\Theta_{\zeta'\zeta'}Z'_iZ'_j + \Theta_{\zeta\zeta'}(Z_iZ'_j + Z'_iZ_j) + \Theta_{\zeta\zeta}Z_iZ_j)\,d\tau\Big) = 0 \qquad (i,j=1,2)$$

notwendig und hinreichend.

Wegen der Symmetrie der Ausgangslösung

$$\zeta = 2m^2\kappa/(ja^2) + \xi\cos j\tau + \cdots = \kappa Z_0 + \xi Z_1(\tau) + \cdots$$

zerfällt die Determinantengleichung (*) in die beiden Faktoren $(i=1,2)$

$$(**) \qquad \kappa\int_0^{2\pi}(2\Theta^{(2)}(Z_iZ_i) + 3.2\Theta^{(3)}(Z_0Z_iZ_i))\,d\tau$$
$$+ \xi^2.4.3\int_0^{2\pi}\Theta^{(4)}(Z_1Z_1;Z_iZ_i)\,d\tau + \cdots = 0.$$

Dabei sind die polarisierten Formen in unserem Fall

$$-\Theta^{(3)}(Z_0ZZ)=(m+1)\left(m(m+1)Z_0Z^2+\tfrac{1}{3}Z_0Z'^2+\tfrac{2}{3}Z'_0ZZ'\right)$$

und

$$\begin{aligned}-\Theta^{(4)}(Z_1Z_1ZZ)&=\tfrac{1}{8}(m+1)^2(17m^2+6m+1)Z_1{}^2Z^2+\tfrac{1}{8}Z'_1{}^2Z'^2\\&\quad-\tfrac{1}{4}(m+1)(-m+5)\left(\tfrac{1}{6}Z_1{}^2Z'^2+\tfrac{1}{6}Z'_1{}^2Z^2+\tfrac{2}{3}Z_1Z'_1ZZ'\right).\end{aligned}$$

Mit Rücksicht auf die Integrale

$$\int_0^{2\pi}\begin{pmatrix}\cos^4\\ \sin^4\end{pmatrix}j\tau d\tau=\tfrac{1}{4}(1+\tfrac{1}{2})2\pi,\qquad\int_0^{2\pi}\cos^2 j\tau\sin^2 j\tau d\tau=\tfrac{1}{4}(1-\tfrac{1}{2})2\pi$$

bekommt man (**) ausgerechnet

$$\begin{aligned}&2\kappa(B_{1020}+j^2B_{1002}+3B_{0030}Z_0+j^2B_{0012}Z_0)\pi\\&\quad=-4.3\xi^2\{(B_{0040}+j^4B_{0004})(2\pm1)+j^2B_{0022}(\tfrac{2}{6}(2\mp1)\pm\tfrac{2}{3})\}\pi/4\\&\quad=-(2\pm1)\xi^2\{3B_{0040}+3(m+1)^4B_{0004}+(m+1)^2B_{0022}\}\pi.\end{aligned}$$

Wegen

$$a^{-2}m^2=(m+1)^{4/3}m^{2/3}>1 \tag{20}$$

ist

$$-\{3B_{0040}/(m+1)^2+B_{0022}+3(m+1)^2B_{0004}\}=\tfrac{1}{2}(14m^2+4m-1)>0,$$
$$m=1,\pm2,\pm3,\cdots$$

sowie

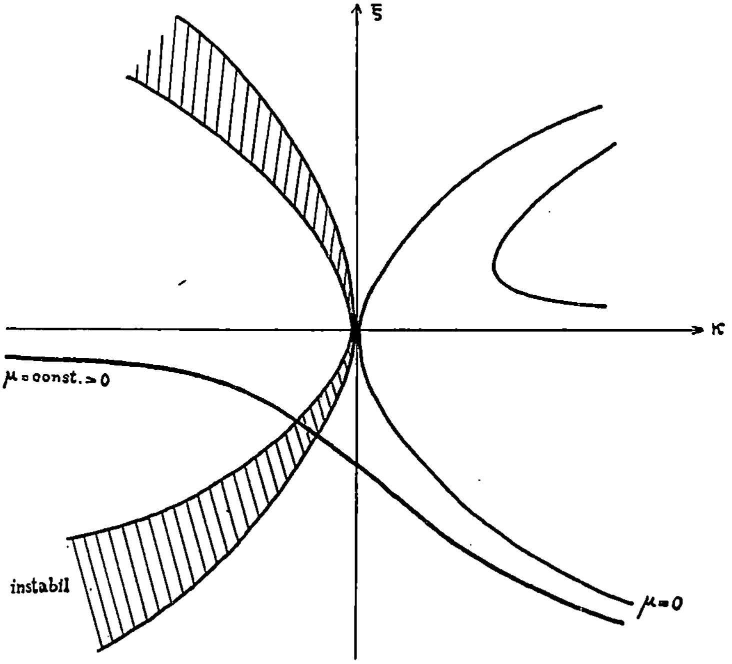
$$\begin{aligned}&-2(B_{1020}+j^2B_{1002}+3B_{0030}Z_0+j^2B_{0012}Z_0)\\&\quad=j(-3(3m+1)-(m+1)+(6m+2m+2)2m^2/a^2)\\&\quad=2(m+1)[(8m+2)m^2/a^2-(5m+2)]\\&\quad=2(m+1)(2(4m+1)(m^2a^{-2}-1)+3m)>0,\end{aligned}$$
$$m=1,\pm2,\pm3,\cdots.$$

Wir bekommen daher endgültig die beiden *parabelartigen Kurven*

$$\begin{aligned}\kappa&=-3\gamma_0\xi^2+\cdots,\\ \kappa&=-\gamma_0\xi^2+\cdots,\end{aligned}\qquad\gamma_0=\frac{m+1}{4}\,\frac{14m^2+4m-1}{2(4m+1)(m^2a^{-2}-1)+3m}>0;$$
$$m=1,\pm2,\pm3,\cdots.$$

Sie sind beide in Richtung abnehmenden κ geöffnet. In dem (schraffierten) Gebiet *zwischen* ihnen sind die periodischen Bahnen einer Gruppe $\mu=\mu_0=\text{const.}$ *instabil*.

Auch die halbganzen Hill'schen Periodenquotienten sind beim restringierten Dreikörperproblem zu beachten. Die Instabilitätsintervalle um diese Stellen herum sind von Wintner [2] untersucht worden. Hier liegt keine Verzweigung der periodischen Bahnen vor. Es gilt mit dem entsprechenden Zuwachs $\kappa = k - k^0$ der Jacobischen Konstante eine Entwicklung der Gestalt



$$\zeta = \kappa Z_0 + \mu Z_* + \cdots .$$

Das Instabilitätsgebiet besteht aus zwei vom Punkt $\kappa = 0$, $\mu = 0$ ausgehenden Winkelräumen $\mu > 0$ und $\mu < 0$, die in erster Annäherung durch zwei „Gerade"

$$\kappa = \gamma_1 \mu + \cdots \text{ und } \kappa = \gamma_2 \mu + \cdots$$

begrenzt werden.

UNIVERSITY OF LEIPZIG, GERMANY.

---

[2] A. Wintner, "On the periodic analytic continuations of the circular orbits in the restricted problem of three bodies," *National Academy of Sciences Proceedings*, vol. 22 (1936), pp. 435-439.

GEODESIC VERTICES ON SURFACES OF CONSTANT CURVATURE.*

By S. B. JACKSON.

1. Introduction. In a previous paper by the writer [6] [1] the attempt was made to characterize structurally, as far as possible, the closed plane curves of class C'' which have exactly two extrema of the curvature; i. e. two vertices. In that paper there were obtained five structural properties. The first part of the present paper is concerned with a discussion of the extent to which these properties can be carried over to characterize closed curves of class C'' which lie in a simply connected region of a surface of constant Gaussian curvature and which have only two geodesic vertices, i. e. extrema of the geodesic curvature. Four of these properties go over without alteration and one important new one is added (Theorem 6. 1) but the fifth one requires certain modifications (Theorems 8. 1 and 9. 1). The essential difference between the case of the surface and that of the plane lies in the geodesic circles which may have different structural properties from plane circles. In particular, they may have more than two points of intersection, a fact which is the basis for some of the examples given in the paper (**9** and **11**).

The last section of the paper (**11**) contains proofs for curves on surfaces of constant curvature of two theorems relating the number of geodesic vertices on a simple closed curve with its number of intersections with a geodesic circle (Theorems 11. 1 and 11. 2). These are direct extensions of known theorems on plane curves [6, Theorems 6. 1 and 7. 1] and are analogues of well known theorems on ovals [3, 1].

Scherk [9] has observed that by stereographic projection any theorem regarding vertices of plane curves is equivalent to one on the sphere concerning geodesic vertices. For the sphere such theorems as Theorem 5. 1, Theorem 11. 1, and Theorem 11. 2 follow trivially from the known results in the plane. Moreover, by using results of Mohrmann [8] regarding inflection points of curves on an ovaloid, these theorems may be readily obtained on the sphere directly, thence yielding simple proofs of the theorems in the

---

* Received March 16, 1948; revised July 19, 1949. Presented to the American Mathematical Society, September, 1947.

[1] The numbers in brackets refer to the bibliography.

plane.[2] It is interesting that these methods of Mohrmann were available in 1917, some years before the first explicit statements in the literature that every simple closed plane curve of class C'' has at least four vertices. As far as this writer is aware, the first statements of this theorem were by Fog [4] in 1933 and Graustein [5] in 1937, and neither of them used Mohrmann's methods. In the present paper however we consider curves in any simply connected region of a surface of constant curvature, without reference to its embedding in space. Mohrmann's methods do not appear directly applicable to this work therefore. Indeed, from one point of view this paper is an answer in one special case to a question raised by Scherk [9] as to what parts of these theorems can be salvaged in case the curves do not lie on an ovaloid.

2. Definitions and previous results.

A *geodesic vertex* of an arc or curve of class C'' on a surface Σ is a point (or an arc of constant geodesic curvature) for which the geodesic curvature has a relative extremum with respect to the neighboring arcs. That is, if $1/\rho$ is the geodesic curvature at any point and $1/a$ the geodesic curvature at the vertex, then in a neighborhood of the vertex $1/\rho - 1/a$ does not change sign and is not identically zero. If, for two geodesic vertices, the geodesic curvatures are both relative maxima or both relative minima, they are said to be of the same type. For a plane curve, geodesic vertices coincide with ordinary vertices.

An arc on which the geodesic curvature is monotone non-increasing or monotone non-decreasing is called a *monotone arc.* If two monotone arcs both have the geodesic curvature non-increasing (non-decreasing) they are said to be of the same type, otherwise of opposite type. An arc or curve for which the geodesic curvature remains constant is called a *geodesic circle.*[3] A geodesic is a special case of a geodesic circle.

In discussing arcs lying on a surface it is often convenient to speak of one arc as lying locally to the right or to the left of another. Such a statement always implies the surface is viewed from the tip of the positive unit normal, and that the surface trihedral is right handed. An arc with positive geodesic curvature at a point thus lies locally to the left of the directed tangent geodesic at this point.

A simple closed arc of a curve which is never crossed by the remainder of the curve is called a *simple loop*. If a simple closed curve, composed of differentiable arcs but allowing corners and lying in a simply connected region of a surface, is directed so that the simply connected region bounded by it lies to its left, we shall say that it is positively directed.

---

[2] Cf. Scherk's review of [6] in *Mathematical Reviews,* vol. 6, p. 100, where he indicates how these theorems may be obtained.

[3] This is what Blaschke [2, § 7] calls a *Krümmungskreis,* as distinguished from an *Entfernungskreis,* which is the locus of points at a fixed geodesic distance from a given point.

A transformation of class C'' of a region $\mathcal{S}$ on a surface Σ into the Gaussian plane is said to be of type I if (a) it is locally one to one, (b) it carries the geodesic circles of $\mathcal{S}$ into circles (or arcs of circles),[4] and (c) it preserves sense. The following results were obtained by the writer in an earlier paper [7, Theorems 3. 1 and 3. 2 and Lemma 4. 7].

THEOREM 2. 1. *A transformation of type I carries monotone arcs into monotone arcs of the same type. It carries geodesic vertices into geodesic vertices of the same type or into limit points of such vertices.*

THEOREM 2. 2. *There is a transformation of type I, not necessarily one-to-one, taking any simply connected region* $\mathcal{S}$ *of a surface* Σ *of constant curvature into a region of the Gaussian plane.*

THEOREM 2. 3. *If* C *is a positively directed simple closed curve, composed of a finite number of arcs of class* C', *which lies in a simply connected region* $\mathcal{S}$ *of a surface of constant curvature* Σ, *the image of* C *under any transformation of type I can contain no simple loop having no points of the curve lying to its left.*

As stated in the previous paper, Theorem 2. 3 was for simple closed curves of class C'', but the proof holds without modification for the more general case.

In this paper we shall be concerned with arcs and curves having continuous geodesic curvature in a simply connected region $\mathcal{S}$ of a surface of constant curvature Σ, and this will be understood in all that follows except where it is indicated otherwise. The region $\mathcal{S}$ is assumed to contain no singularities of any kind. As it is used here, the term *arc* means the locally topological image of a line segment. This differs from the common topological use of the term in that the mapping need not be one-to-one in the large, so an arc may have double points. Similarly a *curve* is the locally topological image of a circle.

3. Geodesic circles. On a general surface, the locus of points at a given geodesic distance from a fixed point, or a distance circle as it is sometimes called, need not be a geodesic circle as defined above. However, it is well known that for surfaces of constant curvature the distance circles are all geodesic circles [2, § 72]. It is not true however that all geodesic circles are distance circles. A distance circle which is a simple closed curve will be called a *complete geodesic circle.*

---

[4] In the Gaussian plane a line is a special case of a circle.

Although it is true, by Theorems 2.1 and 2.2, that geodesic circles in $\mathcal{S}$ go into circles[5] in the plane under transformations of type I, the fact that these transformations are not one-to-one makes it necessary to exercise caution about attributing to geodesic circles in $\mathcal{S}$ many of the familiar properties of circles in the plane. For example, it is not true that two geodesic circles can intersect in only two points. They can have any number of intersections, though under a transformation of type I all the intersections map into one or the other of the intersections of the corresponding plane circles. Examples of geodesic circles which intersect in more than two points will be given in **9** and **11**.

Let $\mathfrak{R}$ denote the closed simply connected subregion of $\mathcal{S}$ bounded by any Jordan curve C in $\mathcal{S}$. Consider the family of complete geodesic circles contained in $\mathfrak{R}$ with centers at a fixed arbitrary point P in $\mathfrak{R}$, and let r be the least upper bound of the radii of these geodesic circles. Since $\mathfrak{R}$ contains no boundary point of $\mathcal{S}$, it is clear that these geodesic circles are complete. If Σ has positive curvature, r must be less than the distance from P to a conjugate point P', for otherwise $\mathcal{S}$ would contain the region covered by all the geodesics through P and hence through P'. Since this region is isometric with a complete sphere, $\mathcal{S}$ would not be simply connected. If the curvature of Σ is non-positive there are no conjugate points in $\mathfrak{R}$. The geodesic circle, O, with radius r and center P has at least one point in common with C, for if it did not, O and C, being closed sets would have a positive distance. A geodesic circle about P of larger radius would then belong to $\mathfrak{R}$, which contradicts the definition of r. Clearly also no point of C lies inside O. From this and the discussion of conjugate points, it follows that every point in and on O is joined to P by a unique geodesic which is the path of minimum distance between these points.

If Q is any point on O and R is any point distinct from P on the geodesic segment PQ, the shortest path from R to any point of O is this geodesic segment RQ. For consider any other arc RS joining R to a point of O. We see that $PR + RQ = r = PS < PR + RS$ whence $RQ < RS$ which proves the contention. If, in particular, Q is a point where O meets C, RQ is the shortest path from any point R of PQ to the curve C. The results of this discussion may be summarized in the following statement.

---

[5] See footnote 4.

LEMMA 3.1. *If a Jordan curve C in $\mathcal{S}$ bounds the closed region $\mathfrak{R}$, about any point P of $\mathfrak{R}$ as center can be drawn a unique complete geodesic circle O contained in $\mathfrak{R}$ and meeting C in one or more points. If Q is one of these common points and R is any point of the geodesic radius PQ, the minimum distance from R to C is along the geodesic segment RQ.*

LEMMA 3.2. *If $\mathfrak{R}$ is the closed region bounded by a Jordan curve C in $\mathcal{S}$, and if C is divided in any manner into three arcs $\mathfrak{a}_1$, $\mathfrak{a}_2$, $\mathfrak{a}_3$, then there exists a complete geodesic circle contained in $\mathfrak{R}$ and having points in common with all three arcs.*

In view of Lemma 3.1, the proof of Lemma 3.2 is an immediate adaptation of one suggested by Paul Erdös for the plane case [6, Lemma 3.1].

LEMMA 3.3. *If a Jordan curve C in $\mathcal{S}$ bounds the closed region $\mathfrak{R}$, and P_0 is any point interior to $\mathfrak{R}$, the largest complete geodesic circle in $\mathfrak{R}$ which contains P_0 either in or on it has at least two points in common with C.*

It is clear that the radius of the largest geodesic circle in $\mathfrak{R}$ with center P is a continuous function of P. Thus the set of points P for which P_0 belongs to the corresponding circle is compact, whence the maximum circle, O, mentioned in the lemma actually exists.

Lemma 3.3 may be made intuitively evident as follows. If O meets C only at a single point Q, a slight displacement of O yields a circle still containing P_0 but having no points in common with C. This contradicts the maximal property of O. A formal proof of the lemma can readily be constructed on these lines.

LEMMA 3.4. *A complete geodesic circle in $\mathcal{S}$ and its interior map in a one-to-one manner into the corresponding plane region under any transformation of type I.*

Under a transformation of type I a complete geodesic circle, being closed, maps into a plane circle traced one or more times. If the plane circle were traced more than once it would contain a simple loop without points of the curve to its left, which contradicts Theorem 2.3. The transformation is thus one-to-one on any complete geodesic circle, or, in fact on any two tangent complete geodesic circles since their images could meet only once. The lemma is now readily proved by showing any two points of the indicated region lie on such a pair of tangent complete geodesic circles.

Let O_1 represent any geodesic circle in $\mathcal{S}$, not necessarily complete. By a transformation of type I it goes into an arc of a circle, possibly overlapping itself. If O_1 has a double point, it must be a point of tangency, since this is true of the plane curve. O_1 is then merely a simple closed curve which maps one-to-one into its plane image, as in Lemma 3. 4. Applying Lemma 3. 2, we find that there is a complete geodesic circle O'_1 interior to O_1 and tangent to it at at least two distinct points. O_1 and O'_1 coincide near these tangencies since their plane images do, and are thus identical. We conclude that O_1 is a complete geodesic circle.

From this discussion, the facts about plane circles, and Lemma 3. 4 the following conclusion may be drawn.

LEMMA 3. 5. *A geodesic circle in $\mathcal{S}$ with a double point is a complete geodesic circle. Two geodesic circles, one of which is complete, either do not meet, or are tangent at just one point, or intersect at just two points.*

The lemma is false if the restriction that one circle be complete is removed. (Cf. **9**).

LEMMA 3. 6. *Every geodesic circle in $\mathcal{S}$ divides $\mathcal{S}$ into exactly two parts.*

If geodesic circle O has a double point, the conclusion follows from Lemma 3. 5 and the Jordan Curve Theorem, so it suffices to consider the case when O is a simple open arc.

A neighborhood of a point of $\mathcal{S}$ bounded by a complete geodesic circle about the point is called a complete circular neighborhood. The closure of such a neighborhood, by Lemma 3. 5, has at most a single arc in common with O, whence no point of $\mathcal{S} - O$ is a limit point of O. Moreover points of such a circular neighborhood which meets O may be classified as locally to the right or left of O. It follows readily that O divides $\mathcal{S}$ into at most two connected sets, namely the sets of points which can be joined in $\mathcal{S} - O$ to points lying locally to the right or left of O respectively.

It remains only to show that no two points P and Q locally on opposite sides of O can be joined by an arc in $\mathcal{S} - O$. Assume such an arc exists. It is clear that this arc may be completed into a simple closed curve by an arc PQ meeting O at a single point T. The region $\mathfrak{D}$ bounded by this curve can be covered, by the Heine-Borel Theorem, by a finite number of complete circular neighborhoods, the closure of each meeting O at most in a single arc. The subarc of O in $\mathfrak{D}$ is thus the sum of a finite number of closed arcs, whence it has an endpoint in $\mathfrak{D}$ since it meets the boundary only at T. Since O has no endpoint in $\mathcal{S}$ this is impossible and the lemma is proved.

As a result of Lemma 3. 6 it is possible to speak, in the large, of the subregion of $\mathcal{S}$ to the left or to the right of any directed geodesic circle.

LEMMA 3.7. *A geodesic circle contained in a closed bounded region of $\mathcal{S}$ is complete.*

The region can be covered by a finite number of complete circular neighborhoods by the Heine-Borel Theorem. Each of these closed neighborhoods has at most a single arc in common with the given circle O, whence O consists of a finite number of closed arcs. Since, as above, O has no endpoint, this is possible only if O has a double point and is therefore complete by Lemma 3.5.

4. Monotone arcs and geodesic vertices. The following two results characterizing geodesic vertices and monotone arcs are known to be true on any surface of sufficient differentiability [7, Corollary 2.1, Lemmas 2.1 and 2.1].

LEMMA 4.1. *A necessary and sufficient condition that an arc be monotone is that it cross every osculating geodesic circle at the point or arc of contact. The geodesic curvature is monotone non-decreasing or monotone non-increasing according as the crossing is from right to left or from left to right.*

LEMMA 4.2. *In some neighborhood of a geodesic vertex, an arc lies entirely to one side of the osculating geodesic circle at this vertex, and conversely, a point (or arc) with this property is a geodesic vertex or a limit point of geodesic vertices. The geodesic curvature at the vertex is a maximum or a minimum according as the arc lies to the right or left of the osculating geodesic circle.*

In both cases the arc and the geodesic circle have locally only a single point or a single arc in common.

LEMMA 4.3. *If P_0 and P_1 are any two points in that order on a monotone arc $\mathcal{A}$ in $\mathcal{S}$, the osculating geodesic circle at P_1 lies to the left or right of that at P_0 according as the geodesic curvature on $\mathcal{A}$ is non-decreasing or non-increasing. The two geodesic circles have no point in common unless they coincide and $\mathcal{A}$ contains the arc P_0P_1 of this geodesic circle.*

Let $\mathcal{S}$ be mapped into the plane by a transformation of type I, which is possible by Theorem 2.2. Since geodesic circles go into circles and monotone arcs into monotone arcs under a transformation of type I, by Theorem 2.1, the fact that the osculating geodesic circles at P_0 and P_1 meet only if they are identical and coincide with $\mathcal{A}$ from P_0 to P_1 follows from the corresponding known result in the plane [6, Lemma 2. 5].[6] This implies immediately that once a monotone arc leaves its circle of curvature it never meets it again. The location of the geodesic circle at P_1 with reference to that at P_0 is then a consequence of Lemma 4. 1.

LEMMA 4. 4. *A monotone arc in $\mathcal{S}$ is simple except for any complete geodesic circles it contains. In this case the arc is tangent to itself without crossing at a single point or along a single arc of such a circle.*

This follows immediately by Lemma 4. 3. Since an arc without a geodesic vertex is monotone, Lemma 4. 4 leads at once to the following.

LEMMA 4. 5. *A simple closed arc in $\mathcal{S}$, not a geodesic circle, has at least one geodesic vertex interior to the arc.*

Let AB be a monotone arc of $\mathcal{S}$, not an arc of a geodesic circle, and let it be tangent to a geodesic circle O_1 at the point of minimum curvature, say B, and lie locally to the left of O_1 at B. Since AB lies locally to the left of O_1 at B, the same is true of its osculating geodesic circle at B. By Lemma 4. 3 A is definitely to the left of this geodesic circle and thus is not on O_1. A similar argument shows that if AB is tangent to and locally to the right of a geodesic circle O_2 at its point of maximum curvature, the other end is not on O_2. Taken together these statements establish the following result.

LEMMA 4. 6. *If an arc AB in $\mathcal{S}$, not an arc of a geodesic circle, is tangent to a geodesic circle in the same direction at A and B and lies locally on the same side of this circle at A and B, then there is a geodesic vertex interior to AB.*

We shall conclude this discussion with a result on plane arcs that will be useful in the next section.

LEMMA 4. 7. *Let a non-circular plane arc AB satisfy the following conditions:*

(a) *it is tangent to a circle (line) in the same direction at A and B;*

(b) *it lies locally to the left of this directed common tangent circle (line) at A and B;*

(c) *it contains no minimum of the curvature in its interior. Then AB*

---

[6] As was mentioned by Scherk in his review of [6] (*Mathematical Reviews*, vol. 6, p. 100) the work on plane monotone arcs is not new in the literature. See, for example, the papers by Vogt and Kneser there mentioned. As a matter of convenience however, references are given to [6] rather than to these original papers.

contains a simple loop having no points of the arc lying to its left. AB meets the circle (line) only at A and B, which may coincide.

Since, by a direct circular transformation, the given circle may be carried into a line, and all the properties are invariant under such transformations, it is enough to consider this case. By Lemma 4. 6, which certainly applies to the plane, AB is not monotone, and therefore has a maximum of curvature, which may be a point or an arc. Let M denote this point, or a point of this arc. As in the proof of Lemma 4. 6, the monotone arcs AM and MB lie entirely to the left of this common tangent line, so AB meets this line only at A and B. The arc AB is not simple since a simple arc satisfying the conditions contains a minimum [6, Cor. 4. 1. 1]. For the same reason it would not be simple even if we deleted from AB all the complete circles noted in Lemma 4. 3, so AM and MB meet. Since AM and MB are monotone and the curvature is non-negative at A and B by condition (b), it is positive interior to AB, and these arcs are respectively inwinding and outwinding spirals [6, Cor. 2. 5. 1 and 2. 5. 2]. The proof that the maximum of curvature lies on a simple loop with no points of AB to its left is identical with the proof that the maximum on a curve with two vertices lies on such a loop [6, Lemma 5. 4]. It will not be repeated here.

5. Location of geodesic vertices.

LEMMA 5. 1. *If a simple arc AB in $\mathcal{S}$, not an arc of a geodesic circle, is tangent to a complete geodesic circle O in the same direction at A and B and never crosses this circle, there is at least one minimum or at least one maximum of the geodesic curvature interior to AB according as AB lies to the left or to the right of this geodesic circle. A and B may coincide.*

If the arc AB lies interior to O, the lemma follows from the one-to-one mapping of type I guaranteed by Lemma 3. 4 and the corresponding known result in the plane [6, Lemma 4. 1]. Let AB then lie outside O. The arc AB may be completed into a simple closed curve C of class C' by adjoining the directed circular arc BA. Let $\mathfrak{R}$ denote the closed region bounded by C. It is sufficient to prove the lemma when AB is so directed that C is a positively directed curve.

Consider first the case when O is exterior to $\mathfrak{R}$, whence AB lies to the left of O. Let C be mapped by a transformation of type I into the plane curve C. Arc AB contains a minimum of the geodesic curvature since otherwise Lemma 4. 7 and Theorem 2. 3 yield contradictory results on its plane image, which establishes the lemma in this case.

In the contrary case O is contained in $\mathfrak{R}$ and the arc AB lies to the right of O. It remains only to show the existence of a maximum on AB. By Lemma 4.6, AB is not monotone, and it is clearly sufficient to consider the case when there is just one geodesic vertex, a point of which we denote by E. Let RS be a subarc of AB containing E and contained in a complete geodesic circle O about E. Consider the complete geodesic circle O'' guaranteed in Lemma 3.2 meeting arcs RE, ES, and $SBAR$ of C. The points of contact are necessarily tangencies. Since O'' clearly cannot meet circular arc BA, by Lemma 4.6 it contains E only if it contains arc RS, in which case this arc is a maximum by Lemma 4.2. If E is not on O'' a subarc of RS is tangent to O'' at two points and the existence of a maximum is assured by mapping O' by a transformation of type I and using Theorem 2.1 and the known results in the plane [6, Lemma 4.1]. This completes the proof of the lemma.

The restriction that AB shall not cross O in Lemma 5.1 may be lightened as follows.

LEMMA 5.2. *If a simple arc AB in $\mathcal{S}$, not an arc of a geodesic circle, is tangent to a complete geodesic circle O in the same direction at A and B and lies locally on the same side of O at A and B, there is at least one minimum or at least one maximum of the geodesic curvature interior to AB according as AB lies locally to the left or right of O at these points. A and B may coincide.*

By Lemma 4.6 there is at least one geodesic vertex interior to AB. Consider the case when AB lies locally to the right of O and assume the lemma is false, i. e. assume the only geodesic vertex is a minimum. Since O lies to the left of AB at A and B, it lies to the left of the osculating geodesic circles at A and B. If the only vertex is a minimum, then by Lemma 4.3 O lies to the left of all the osculating geodesic circles. Arc AB thus never crosses O and lies entirely to its right. But Lemma 5.1 then states that AB has a maximum of geodesic curvature. The contradiction proves the lemma for this case. The remaining case when AB is locally to the left of O may be reduced to the one above by reversing the sense on AB. This completes the proof.

Lemma 5.1 leads easily to the following useful result.

LEMMA 5.3. *Let AB be a simple closed arc in $\mathcal{S}$, not a geodesic circle. Let $\mathfrak{R}$ denote the region bounded by AB, and θ the positive angle interior to $\mathfrak{R}$ between the two tangents at the double point. If $\theta \geqq \pi$ there is a maximum or a minimum of the geodesic curvature interior to AB according* *as $\mathfrak{R}$ lies to the right or left of AB. If $\theta \leqq \pi$ there is a maximum or a minimum of the geodesic curvature interior to AB according as $\mathfrak{R}$ lies to the left or right of AB.*

It is clearly possible to draw an arbitrarily small complete geodesic circle tangent to AB at points A' and B' near A and B. It may be that these four points will coincide. This complete geodesic circle can be drawn exterior to $\mathfrak{R}$ or interior to $\mathfrak{R}$ according as $\theta \geqq \pi$ or $\theta \leqq \pi$. The proof is then immediate from Lemma 5. 1.

It is interesting to observe that Lemma 5. 3 remains valid when $\mathfrak{R}$ is thought of as the region *exterior* to AB in $\mathcal{S}$. This follows easily from the lemma itself.

Lemma 5. 3 leads trivially to one of the main results of a previous paper [7, Theorem 4. 1], namely the Four-vertex Theorem for $\mathcal{S}$.

THEOREM 5. 1. *Every simple closed curve of class C'', not a geodesic circle, in a simply connected region $\mathcal{S}$ of a surface of constant curvature has at least four geodesic vertices.*

Let the curve be positively directed, and let M be a geodesic vertex. Consider the arc from M to M as the arc AB of Lemma 5. 3. Since $\theta = \pi$ the arc contains both a maximum and a minimum distinct from M. This proves the theorem since the number of geodesic vertices is even if it is finite.

6. Curves with two geodesic vertices. We shall turn our attention to obtaining structural properties of curves in $\mathcal{S}$ having exactly two geodesic vertices. Such a curve, C, consists of two monotone arcs of opposite type. As was noted in Lemma 4. 4, these monotone arcs may contain complete geodesic circles, and one or both of the geodesic vertices may be such circles. The curve which is obtained from C by removing all such complete geodesic circles is called the normalized curve, $\bar{C}$, corresponding to C. $\bar{C}$ is still of class C'', and the process neither gains nor loses geodesic vertices.[7] In the following discussion $\bar{C}$ always denotes the normalized curve corresponding to C.

A double point is called simple if the curve passes through the point only twice.

We proceed to establish the following five properties of C.

THEOREM 6. 1. *If C is a curve of class C'', not a geodesic circle, in a simply connected region $\mathcal{S}$ of a surface of constant curvature, and if C has exactly two geodesic vertices, then:*

---

[7] If a complete geodesic circle is traced completely k times and partly so again, it is understood that in $\bar{C}$ the k complete revolutions are omitted, but the remaining arc left, so $\bar{C}$ is closed and of class C''.

(a) *the normalized curve* $\bar{C}$ *may be divided into two simple arcs;*

(b) *the normalized curve* $\bar{C}$ *has double points, but all of them are simple;*

(c) *at any point of tangency of* C *with itself the directed tangents coincide;*

(d) C *contains exactly two simple loops, one loop containing the maximum of geodesic curvature and having no points of* C *to its left, the other containing the minimum of geodesic curvature and having no points of* C *to its right;*

(e) C *has only a finite number of double points and double arcs.*

Property (a) follows from Lemma 4.4 and the fact that the two monotone arcs making up $\bar{C}$ are simple by construction.

In (b), the fact the $\bar{C}$ has double points is precisely the content of Theorem 5.1. If P were a double point of $\bar{C}$ which was not simple, it would divide $\bar{C}$ into at least three closed arcs, none of which are geodesic circles by construction of $\bar{C}$. By Lemma 4.5 each of these arcs would have a geodesic vertex, contradicting the hypothesis of only two geodesic vertices on C. Property (b) can also be proved directly from Lemma 4.4, since neither of the two monotone arcs can pass through any point but once.

To prove (c) let $\mathcal{S}$ be mapped into the plane by a transformation of type I. The result then follows from the corresponding property for plane curves with just two vertices [6, Theorem 5.1].

For the proof of (d), let m and M be the points of minimum and maximum geodesic curvature respectively (or points on the arcs of minimum and maximum geodesic curvature). If the maximum of geodesic curvature is a complete geodesic circle, this circle is itself the required simple loop since, by Lemma 4.3, no point of C lies to its left. In the contrary case, by Lemma 4.4, arcs mM and Mm are simple except for any complete geodesic circles they may contain. Moreover they intersect, for otherwise C is simple, contrary to (b). Let A be the first point where Mm meets mM. We shall show that AMA is the required simple loop. The two arcs are not tangent at A, for the directed tangents would coincide by (c), whence the arc would contain both a maximum and a minimum by Lemma 5.3, and this is impossible. Since M is a maximum, by Lemma 5.3 the arcs pass at A into the region to the right of AMA. It should be noted that this region may be the exterior of the arc. By construction, arc mA does not cross AMA and thus lies to its right. If AMA is not the required simple loop the arc Am must cross it. Let B be the first such crossing. Since Mm never crosses itself, B must lie on AM, and if B is the first crossing Am approaches

AM from the right at B. Consider arc $BMAB$. If the region $\mathcal{R}$ bounded by it (inside it) lies to its left the angle θ interior to $\mathcal{R}$ between the tangents at B is greater than or equal to π. If $\mathcal{R}$ lies to the right of $BMAB$, then $\theta \leqq \pi$. In either case there would be a minimum on $BMAB$ by Lemma 5. 3. Since this is false, the point B does not exist and AMA is the required simple loop. The case of the minimum of geodesic curvature is reduced to this by reversing the direction on C, which completes the proof of (d).

Finally, consider property (e). Whenever one of the two monotone arcs meets itself, it contains a complete geodesic circle by Lemma 4. 4. These double points and arcs are thus surely finite in number, and hence isolated on the arc. At a point where the two monotone arcs are tangent to each other but have different geodesic curvatures, the directed tangents coincide, by (c), and it is readily verified that the arcs have locally only the single point of tangency in common, so that such tangencies are isolated double points. At a point or arc of tangency where the geodesic curvatures are equal, the osculating geodesic circles coincide. Indeed the arc of contact, if any, is an arc of this circle since one arc has monotone non-decreasing and the other monotone non-decreasing geodesic curvature. But by Lemma 4. 1 one of the arcs crosses this circle from right to left and the other from left to right. The osculating geodesic circle thus acts as a barrier and the two arcs have no other points in common in some neighborhood of this point or arc of contact which is thus isolated. If C had an infinite number of double points and arcs, they would have a limit point, which would necessarily be a tangency. But since all the tangencies have been shown to be isolated this is impossible, and property (e) is established. Properties (a), (b), (c), (d) above are already known for plane curves with two vertices [6, Theorem 5. 1], while property (e) is new.

By (e) the curve C divides $\mathcal{S}$ into a finite number of regions, which are all simply connected except one, called the exterior.[8] Two of the regions, by (d), are completely bounded by the simple loops. For plane curves it is known [6, Theorem 5. 1] that *no region determined by* C *except the two indicated in* (d) *is bounded in the same sense by all the arcs of* C *which bound it.*

The next three sections of the paper are devoted to a discussion of the validity of this statement for curves in $\mathcal{S}$.

---

[8] In the case of the Gaussian plane, or equivalently the sphere, there is no real distinction between interior and exterior regions, since any finite point can be carried into the point at infinity by a direct circular transformation. In general, however, the distinction between exterior and interior regions is genuine, as will be made clear in the work that follows.

7. Angular measure of transformed curves. Let C be a positively directed simple closed curve of class[9] D' in $\mathcal{S}$, and let K be the plane image of C under a transformation of type I which carries the interior of C into a finite region of the plane. Such a transformation always exists unless every point of the plane is the image of some point in this interior. K is also of class D'. Let C be deformed, as is always possible, through simple closed curves of class D' lying in its interior into a simple closed curve C_0 lying in a complete geodesic circle of $\mathcal{S}$, and let the deformation be such that at any corner the angle between the directed tangents remains in the open interval $(-\pi, \pi)$. Curve K is thereby similarly deformed through finite points into a curve K_0, which is a positively directed simple closed curve by Lemma 3. 4.

The angular measure of a closed plane curve of class D' is defined as the total rotation of a directed tangent on tracing the curve once. At a corner this includes the directed angle in the interval $(-\pi, \pi)$ through which the first directed tangent must turn to coincide with the second. This is equivalent to rounding off each corner with a small circular arc and considering the angular measure of the resulting curve of class C'. Since K_0 is a simple closed curve, its angular measure is 2π, [10]. We conclude K also has angular measure 2π since the angular measure varies continuously (the deformation being through finite points) but is an integral multiple of 2π. This justifies the following statement.

LEMMA 7. 1. *The image, under a transformation of type I, of any positively directed simple closed curve of class D' in $\mathcal{S}$ whose interior has a finite image has an angular measure of 2π.*

One further result regarding the angular measure of plane curves will be convenient.

LEMMA 7. 2. *Let K be a closed plane curve of class D' for which there is a point O not on K, satisfying the following condition: the angle θ from some fixed direction to OP changes monotonically as P traces K. Then the angular measure of K equals the total variation in θ.*

It is sufficient to prove the result when θ is monotone increasing. At a corner of K, the various positions assumed by a directed line rotating from the first directed tangent to the second will also be called tangents, for convenience. Consider a point P of K and a directed tangent at P. The angle θ from the given fixed direction to OP, the angle λ from this same fixed direction to the directed tangent, and the angle ϕ from OP to the directed tangent are clearly related by the equation $\lambda = \phi + \theta$, whence it follows that $\Delta\lambda = \Delta\phi + \Delta\theta$ where the Δ indicates the total variation around K. Since θ is increasing it follows that $0 < \phi < \pi$. But since $\Delta\phi$ is an integral multiple of 2π, it follows that $\Delta\phi = 0$. This proves the lemma, for $\Delta\lambda$ is the angular measure of K.

---

[9] A curve or arc is said to be of class D' if it is a finite succession of differentiable arcs. It is of class C' except for a finite number of corners.

8. Arcs bounding the exterior region.

LEMMA 8. 1. *Let AB and BA be two simple monotone arcs of opposite type in $\mathcal{S}$, meeting each other only at A and B and not tangent to each other in opposite directions at these points. Then there exists a transformation of type I of the closed region $\mathfrak{R}$ bounded by the two arcs which is one-to-one and which carries this region into a finite region of the plane.*

Let the arcs be directed so that the region bounded by them lies to their left. Since the two arcs are of opposite type, one of the ends, say A, is the point of minimum geodesic curvature on both arcs.

Suppose first that the angle interior to $\mathfrak{R}$ at A is less than or equal to π, and let P be any point interior to $\mathfrak{R}$. By Lemma 3. 3 there is a complete geodesic circle containing P which is contained in $\mathfrak{R}$ and has at least two points in common with its boundary. Since B is the only point where this circle could meet the boundary and not be tangent to it, this circle is tangent to one of the two arcs. There exists a complete geodesic circle passing through P which is contained in the one just constructed and has the same point of contact with one of the given arcs. This fact follows from the possibility of the construction in the plane and the fact that a transformation of type I on the complete geodesic circle is one-to-one. The point P therefore lies on a complete geodesic circle tangent to one of the arcs and lying to the left of both the arc and its osculating geodesic circle at the point of contact.

If, on the other hand, the angle at A in $\mathfrak{R}$ exceeds π, draw the osculating geodesic circle to BA at A, which clearly passes to the interior of $\mathfrak{R}$ at A. Since by Lemma 3. 6, it divides $\mathfrak{R}$ into two parts, it meets the boundary again for the first time in a point Q. The point Q must be on arc AB, for BA has no further points in common with this circular arc unless BA is a geodesic circle, in which case Q coincides with B. This circular arc, AQ, divides $\mathfrak{R}$ into the subregions $\mathfrak{R}_1$ and $\mathfrak{R}_2$ where $\mathfrak{R}_1$ denotes the subregion containing arc BA in its boundary. Since the only angle interior to $\mathfrak{R}$ which may exceed π is at B we apply Lemma 3. 3 as before to any point P of $\mathfrak{R}_1$, and again conclude that P lies on a complete geodesic circle which is tangent to and to the left of one of the osculating geodesic circles of the given arcs. If P belongs to $\mathcal{R}_2$, Lemma 3. 3 assures us of a complete geodesic circle containing P and meeting the boundary of $\mathcal{R}_2$ at least twice. The angles interior to $\mathcal{R}_2$ at A and Q are acute, so both contacts are tangencies. Moreover, by Lemma 3. 5 a geodesic circle cannot be tangent to a complete geodesic circle at two points, so one of these tangencies is on the given arc AQ. The argument used above shows as before that P lies on a complete geodesic circle tangent to and to the left of an osculating geodesic circle of one of the given arcs. This property has now been established for all points in $\mathcal{R}$.

Since the arcs do not have oppositely directed common tangents at A, there exists a point of $\mathcal{S}$ lying to the right of the two osculating geodesic circles at A and contained in a complete geodesic circle about A. Consider any transformation of type I of $\mathcal{S}$ which carries this point into the point at infinity in the plane [cf. 7, Lemma 3. 5]. The two osculating geodesic circles at A are thereby carried into plane circles whose interiors lie to their left. That is so say, the transformed arcs $A'B'$ and $B'A'$ both have positive curvature at A', and since at A' both arcs have their minimum curvature, it follows that both arcs have positive curvature at every point. All osculating circles lie interior to one or the other of the osculating circles at A' by Lemma 4. 3, and all these osculating circles have their interiors to their left. Since we have shown that every point P in $\mathcal{R}$ lies on a complete geodesic circle tangent to and to the left of some osculating geodesic circle of AB or BA, the image point P' lies on a circle tangent to and inside of one of the osculating circles of $A'B'$ or $B'A'$. This shows that the transformation takes every point P of $\mathcal{R}$ into a finite point. In fact the entire region $\mathcal{R}$ is mapped into the sum of the two osculating circles at A'.

By Lemma 7. 1 the curve of class D' consisting of the arcs $A'B'$ and $B'A'$ has an angular measure of 2π. Let this curve be taken as the curve K of Lemma 7. 2. Any point lying to the left of both the osculating circles at B satisfies the conditions of the point O of the lemma, since it lies interior to all the osculating circles, so the vector OP always turns in the positive sense as P traces K. By Lemma 7. 2 $\triangle\theta = 2\pi$; i. e. the radius vector OP turns around exactly once, turning always in the positive sense. This means that K is a positively directed simple closed curve. The transformation from the curve in S to K is then one-to-one. It is known that the interior of K is the topological image of one or more simply connected subregions of $\mathcal{R}$ [7, Lemma 4. 5]. However, since the boundary of $\mathcal{R}$ maps into K one-to-one, the points of $\mathcal{R}$ near its boundary map into the points in the plane close to K and interior to it. Since this set of points of R is connected, they all belong to the same component of the inverse image of the interior of K. But the only simply connected subregion of $\mathfrak{R}$ containing all these points is the entire interior of $\mathfrak{R}$. The transformation is thus proved to be one-to-one on $\mathfrak{R}$, which completes the proof of the lemma.

Lemma 8. 1 gives rise easily to the facts stated in the next two lemmas.

LEMMA 8. 2. *If C is a curve in $\mathcal{S}$ having just two geodesic vertices, and if the exterior region is bounded entirely by one of the simple loops of C, then none of the regions of $\mathcal{S}$ determined by C is bounded in the same sense by all of its bounding arcs except the two bounded entirely by the simple loops.*

LEMMA 8. 3. *If C is a curve in $\mathcal{S}$ having just two geodesic vertices, it is not possible for the boundary of the exterior region to consist of exactly two arcs which bound this region in the same sense.*

For Lemma 8. 2 the arcs AB and BA are the arcs into which the vertex divides the simple loop, while for Lemma 8. 3 they are the two alleged bounding arcs. In both cases Lemma 8. 1 guarantees a one-to-one transformation of type I of the interiors of these arcs, from which the conclusions follow from the known facts for plane curves with two vertices [6, Theorem 5. 1 (d); see also Section **6** above].

We turn finally to the possibility that the boundary of the exterior region for a curve C in $\mathcal{S}$ having just two geodesic vertices shall consist of more than two arcs, all bounding it in the same sense. Assume this to be the case, and let C be so directed that the exterior lies to the right of its bounding arcs. It can be easily verified by use of Theorem 6. 1 that this sequence of arcs, finite in number, form a simple closed curve which is the positively directed boundary of a simply connected region $\mathfrak{R}$.

The arcs of monotone increasing and decreasing geodesic curvature constituting C will be denoted by A_i and A_d, and their subarcs called i-arcs and d-arcs respectively. Since neither A_i or A_d crosses itself, it is clear that the i-arcs and d-arcs alternate on the boundary of $\mathfrak{R}$. Thus if P_1 denotes the point of the boundary of $\mathfrak{R}$ nearest the minimum of geodesic curvature on A_i and if succeeding endpoints taken in order about the boundary of $\mathfrak{R}$ are $Q_1, P_2, Q_2, \cdots, P_n, Q_n$, arcs P_kQ_k are i-arcs while arcs Q_kP_{k-1} are d-arcs. It is assumed that all subscripts are reduced mod n, and by Lemma 8. 3 $n \geqq 2$.

It is relatively simple deduction, using Lemma 4. 4, that the order in which these endpoints occur on A_i and A_d coincides with the cyclic order, given above, in which they occur on the boundary of $\mathfrak{R}$. The first of these points on A_i is P_1, while the first on A_d is, say, Q_s. For every pair of consecutive endpoints except P_sQ_s and Q_nP_1 it follows that there is both an

i-arc and a d-arc joining them. If P is any point of $\mathcal{R}$ between such a pair of arcs, Lemma 3. 3 guarantees the existence of a geodesic circle containing P and tangent to both arcs, since neither arc can be tangent to the circle twice by Lemma 4. 6. As in Lemma 8. 1 we can thus obtain a geodesic circle passing through P, tangent to the boundary of $\mathcal{R}$ and lying to the left of the osculating geodesic circle at the point of contact.

Any other point of R lies to the left of all the d-arcs and i-arcs here mentioned. Since a discussion of the various possible cases, namely $s = 1$, $s = n$, and $s \neq 1, n$, reveals that the subregion of $\mathcal{R}$ bounded by these arcs and containing P can contain at most one interior angle exceeding π, it follows as above using Lemma 3. 3 that there is a complete geodesic circle passing through P, tangent to A_i or A_d and lying to the left of its osculating geodesic circle at the point of contact. This property has now been established for all points of $\mathcal{R}$.

Let $\mathcal{S}$ be mapped into the plane by a transformation of type I taking some point to the right of the osculating geodesic circle of C at the point of minimum geodesic curvature into the point at infinity in the plane. The curvature of the transformed curve K is then always positive, and all osculating circles lie interior to and to the left of the one of minimum curvature, by Lemma 4. 3. By use of Lemma 7. 1 and 7. 2, taking O as any point in the smallest osculating circle of K, it follows exactly as in Lemma 8. 1 that the transformation maps $\mathcal{R}$ one-to-one into a finite portion of the plane. This would mean that the exterior region for the plane curve K has a boundary which consists of more than two arcs all bounding it in the same sense, which is known to be impossible [6, Theorem 5. 1 (d)]. It is therefore impossible for the exterior region of C on $\mathcal{S}$ to be bounded in the same sense by more than two arcs. This result and the facts stated in Lemmas 8. 2 and 8. 3 may be collected in the following theorem.

THEOREM 8. 1. *If C is a curve of class C'', not a geodesic circle, in a simply connected region $\mathcal{S}$ of a surface of constant curvature, and if C has exactly two geodesic vertices, then the exterior region of C on $\mathcal{S}$ can be bounded in the same sense by all its bounding arcs only if it is one of the two regions completely bounded by a simple loop. In this case, no region determined by C except these two is bounded in the same sense by all its bounding arcs.*

9. Arcs bounding an interior region. Theorem 8. 1 answers affirmatively for the exterior region the question as to whether the conjecture at the end of section **6** is valid on $\mathcal{S}$. It remains to consider the interior regions.

LEMMA 9. 1. *If a curve C in $\mathcal{S}$ has just two geodesic vertices, no interior region determined by C whose boundary consists of more than two arcs can be bounded in the same sense by all these bounding arcs.*

Suppose such a region $\mathfrak{R}$ exists. Let any two adjacent arcs of its boundary be called $\mathfrak{C}_1$ and $\mathfrak{C}_2$, and let the remainder of the boundary be called $\mathfrak{C}_3$. By Lemma 3. 2 there is a complete geodesic circle in $\mathfrak{R}$ having points in common with all three arcs. Since all angles interior to $\mathfrak{R}$ at its corners do not exceed π, these contacts are all similarly directed tangencies. The tangencies divide C into at least three arcs, each of which either contains a geodesic vertex by Lemma 4. 6 or is a geodesic vertex by Lemma 4. 2. This contradicts the hypothesis that C has only two geodesic vertices, whence no such region $\mathfrak{R}$ can exist.

The conjecture of section **6** would be completely proved if we could show finally that no interior region could be bounded in the same sense by just two arcs. Strangely enough, this is false, as will be demonstrated by an example.

It is readily shown that the lemniscate, whose polar equation is $r^2 = \operatorname{Cos} 2\theta$, has just two vertices, and that the inflectional tangents at the double point are the osculating circles there. The curve is directed so that the right hand loop is positively traced, inducing a direction also on the inflectional tangents. Certain points on the curve and its tangents will be denoted as follows: $P'(0,0)$, $A'(1,0)$, $B'(-1,-1)$, $D'(1,1)$, $E'(-1,0)$, $F'(1,-1)$, $G'(-1,1)$, and C' the point at infinity, the given coordinates being rectangular. Let us subject the figure to the complex linear transformation $w = (1-iz)/(z-i)$ and denote the transforms of $A', B', \cdots$, by $A, B, \cdots$. In the transformed figure the origin O is exterior to both the loops PAP and PEP but is interior to the two osculating circles at P, which are oppositely directed. Consider the curve $PAPBCDPEPFCGP$ obtained by inserting the osculating circles into the original curve. Since the transformation is of type I the only vertices on this curve are at A and E. Instead of considering this as a curve in the plane let us consider it on the logarithmic Riemann surface with branch point at O, i. e. the Riemann surface for $w = \log z$.

Tracing out the loop PAP brings us back to the same point P of this surface, since O is not contained in this loop, but when we trace the first osculating circle, that does contain O, and therefore leads to a point $\bar{P}$ corresponding to P but on a different sheet of the surface. Upon tracing the second loop, not containing O, we return to $\bar{P}$, but on tracing the other osculating circle, O is traversed in the reverse direction and we therefore return to P. The curve is therefore a closed curve on the Riemann surface. Moreover the osculating circles clearly have a common point C on the surface. On the surface the curve can thus be described as the curve $PAPBCD\bar{P}E\bar{P}FCGP$. The only double points are P, $\bar{P}$, and C, and they are all simple. It is immediately clear that there are two interior regions of this curve which are bounded in the same sense by the two arcs which bound them. They are the region bounded by circular arcs $CD\bar{P}$ and $\bar{P}FC$ and the region bounded by circular arcs PBC and CGP.

Considering the curve as a curve of the Riemann surface was done merely for convenience in the discussion. The configuration can occur on an ordinary developable surface. For example, the conical surface with vertex at the origin and a circular helix about the z-axis as directing curve will serve the purpose, as will the tangent surface of a circular helix. The conjecture of section **6** is thus false for the interior regions of the curve C. It should be emphasized that while, as here, the surface Σ may contain branch points or other singularities, region $\mathcal{S}$ is assumed free of them. Here $\mathcal{S}$ may consist of that part of the Riemann surface covering an annular ring about O. This region is simply connected and free of singularities. Similar remarks apply to the regions $\mathcal{S}$ in the examples of section **11**. The results for interior regions may be summarized in the following theorem.

THEOREM 9.1. *If C is a curve of class C'', not a geodesic circle, in a simply connected region $\mathcal{S}$ of a surface Σ of constant curvature, and if C has exactly two geodesic vertices, any interior region determined by C and bounded in the same sense by all its bounding arcs is either completely bounded by one of the simple loops, or is bounded by exactly two arcs. If regions of this latter type occur, the regions bounded by the simple loops are both interior regions.*

It is interesting to observe that in the example just given the critical point is that the two circles intersect in three distinct points, namely P, $\bar{P}$, and C. By continuing the subarcs so that their plane images are traced n times, these arcs meet in $2n-1$ points and the number of regions bounded as above by two arcs is $2n$. Thus we can construct examples in which the number of regions bounded by two arcs in the same sense is arbitrarily large. As far as the circles themselves are concerned, it should be noted that in the simply connected region consisting of the part of the surface covering a circular ring in the plane about the origin, the circles intersect infinitely many times. Similarly we can construct examples of circles in such a simply connected region which are tangent at infinitely many distinct points.

10. Counterexample. As has already been noted in Lemma 4. 3, monotone arcs have an essentially spiral character. In fact any plane curve with just two vertices can be reduced by a direct circular transformation to a curve in which one of the two monotone arcs is an inwinding spiral, and the other an outwinding spiral [6, p. 573]. For convenience we will suppose the curves normalized, so the only double points are where the two monotone arcs meet each other. The following question then naturally arises. Let the double points or arcs be arranged in the order $P_1, P_2, \cdots, P_n$ in which they occur on one of the monotone arcs. Will they occur in the reverse order $P_n, P_{n-1}, \cdots, P_1$ on the other arc? The same question can be phrased in a different but equivalent way as follows. Is every double point or arc a cut point or arc for the set of points making up the curve C? Simple examples show that this is often the case. If the question could be answered in the affirmative, it would be a very strong structural characteristic of curves with just two vertices. Actually the answer is in the negative, however, as is shown by the following example of a normalized curve in the plane having five double points, for which the order of the points on one arc is P_1, P_2, P_3, P_4, P_5, while on the other arc the order is P_5, P_2, P_3, P_4, P_1.

A curve is determined by the following equations if the radius of curvature R is given as a function of the slope angle ϕ

$$x = x_0 + \int_0^\phi R \operatorname{Cos} \phi \, d\phi \qquad y = y_0 + \int_0^\phi R \operatorname{Sin} \phi \, d\phi.$$

Let $x_0 = y_0 = 0$, and consider the curve (Figure 1) defined as follows for $|\phi| \leqq 4\pi$.

$$R = \begin{cases} 1 + |\phi|, & 0 \leqq |\phi| \leqq \pi \\ 1 + \pi, & \pi \leqq |\phi| \leqq 2\pi \\ 1 + |\phi| - \pi, & 2\pi \leqq |\phi| \leqq 3\pi \\ 1 + 2|\phi| - 4\pi, & 3\pi \leqq |\phi| \leqq 4\pi. \end{cases}$$

As defined above, R is positive and continuous, and the curve consists of two monotone arcs. The arc $-4\pi \leqq \phi \leqq 0$ is an arc of non-decreasing curvature, while the arc $0 \leqq \phi \leqq 4\pi$ is an arc of non-increasing curvature. It is a routine matter to carry out the computation to find the various points where the tangents are horizontal.

These points, together with the corresponding values of ϕ are as follows.

$\phi = 0,$	$O(0, 0)$		
$\phi = \pi$	$A(-2, 2 + \pi)$;	$\phi = -\pi$	$A'(2, 2 + \pi)$
$\phi = 2\pi$	$B(-2, -\pi)$;	$\phi = -2\pi$	$B'(2, -\pi)$
$\phi = 3\pi$	$C(-4, 2 + 2\pi)$;	$\phi = -3\pi$	$C'(4, 2 + 2\pi)$
$\phi = 4\pi$ and -4π,		$D(0, -4\pi)$.	

Since at D the tangents and curvatures of the two arcs coincide, it is clear that this is actually a closed curve of class C'', having just two vertices, O and D, and with angular measure 8π.

Since, by its definition, the curve is symmetric in the y-axis, all the points where the curve meets this axis are double points. P_1 is the first point where arc OD meets the y-axis, and thus, by symmetry, the first double point on this arc. The open subarc P_1AB clearly does not meet the y-axis, but it is readily found that BC crosses the axis twice since B and C have negative x-coordinates, while the point where $\phi = 5\pi/2$ has a positive value for x. These two double points are denoted by P_3 and P_5. However, it appears at once that arc OD has positive slope at P_3, while DO has negative slope there. The crossing is therefore in the direction indicated, proving the existence of the other two symmetric double points P_2 and P_4 on arcs P_1P_3 and P_3P_1 respectively. The double points therefore occur on the monotone arcs in the orders indicated above. This answers in the negative the question asked at the beginning of this section.

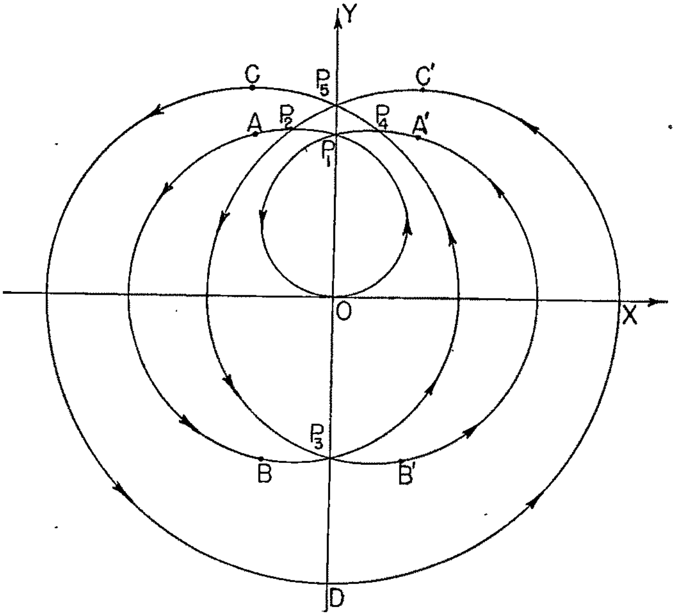


FIGURE 1.

11. Geodesic vertices on simple closed curves. Closely related to the classical four-vertex theorem are several other theorems relating the number of vertices on an oval to the number of its intersections with a circle [1, 2 p. 49, 3]. These theorems can be extended to simple closed plane curves

[6, § 6 and § 7]. It is the purpose of this section to investigate the extent to which these theorems can be generalized to simple closed curves on $\mathcal{S}$.

THEOREM 11. 1. *A simple closed curve* C *of class* C'' *in* $\mathcal{S}$ *not a geodesic circle which meets any geodesic circle at most four times has exactly four geodesic vertices.*

Let C be positively directed and consider a point or arc M of maximum geodesic curvature. We will show that the osculating geodesic circle O at M is a complete geodesic circle lying in the closed region $\mathfrak{R}$ bounded by C and meeting C only at M. The osculating geodesic circle O at M lies locally interior to C near M by Lemma 4. 2. If it is not contained in R, both branches of the circle from M must cross C, since by Lemma 3. 6 each arc of the circle divides R into two parts. It will then be possible to select points A, B, M, B', A' in that order on O with A, A' exterior to C and B, B' interior to C. Let O be deformed into an arbitrarily near geodesic circle by decreasing the geodesic curvature slightly, preserving the tangency with C at M (or some point of M). The points A, A', B, B' will vary continuously into points A_1, A'_1, B_1, B'_1 on the new circle O_1, but the deformation may be taken so small that these points do not cross C. The new circle lies locally to the right of C at M by Lemma 4. 2, so it is possible to choose points P, P' of O_1 exterior to C on arcs MB_1 and MB'_1 respectively. Thus, in addition to M, C meets O_1 on each of the arcs A_1B_1, B_1P, $P'B'_1$, $B'_1A'_1$, which contradicts the assumption of at most four intersections. If O is contained in $\mathfrak{R}$ it is necessarily a complete geodesic circle by Lemma 3. 7. If it meets C in some point of C other than M, this is a point of tangency. The deformation of O discussed above will take this point outside of C and yield the same contradiction as before. It follows that O is a complete geodesic circle in $\mathfrak{R}$ meeting C only at the vertex M.

The remainder of the proof of this theorem, based on Lemma 3. 2, is identical with that for the case of the plane [6, Theorem 6. 1], and need not be repeated.

THEOREM 11. 2. *Let a simple closed curve* C *of class* C'' *in* $\mathcal{S}$, *not a geodesic circle, be met by a complete geodesic circle* O. *If, among the arcs into which* O *divides* C, *there can be found* n *arcs* $P_{2i-1}P_{2i}$ $(i = 1, 2, \cdots, n)$ *interior to* O *such that the points* P_k *are in the same cyclic order on* C *and* O, *then* C *has at least* $2n$ *geodesic vertices.*

It is possible to restrict attention entirely to the case when none of the P_k are points of tangency, since in any case this can be arranged by a slight deformation of O. It should be noted that the theorem does not require that the arcs selected be all the arcs of C interior to O.

The proof of this theorem is identical with the proof of the corresponding result in the plane [6, Theorem 7. 1] except that in the present case we have assumed outright the existence of the n interior arcs. For the present proof Lemma 3. 2 and Lemma 5. 2 replace Lemma 3. 1 and Corollary 4. 1. 1 respectively of the former paper. The proof will not be duplicated here. The following immediate corollary is easier to visualize, though somewhat less general.

COROLLARY 11. 2. 1. *If a simple closed curve C of class C'' in $\mathcal{S}$, not a geodesic circle, intersects a complete geodesic circle O in just $2n$ points, and if these intersections have the same cyclic order on C and O, then C has at least $2n$ geodesic vertices.*

In comparing this last theorem with the corresponding results in the plane, at least two questions naturally arise, as follows.

(a) Is Theorem 11. 2 still true if O is taken as any geodesic circle, rather than a complete geodesic circle?

(b) The proof of Theorem 11. 2 is based essentially on Lemma 5. 2 which assures us of the existence of a certain type of vertex on an arc AB tangent to a complete geodesic circle O. Is this lemma itself true if O is any geodesic circle rather than a complete geodesic circle?

We shall proceed to show by counterexamples that both questions are to be answered in the negative. In other words, the restrictions indicated in the theorem above are essential. For convenience the examples will be constructed on the Riemann surface of **9**, but as noted there, they can be realized on an ordinary surface.

Consider any ellipse $x^2/a^2 + y^2/b^2 = 1$ whose eccentricity is greater than $1/\sqrt{2}$, so that the osculating circles at the points $(\pm a, 0)$ do not intersect. Let this figure be subjected to a direct circular transformation taking the center of curvature at $(-a, 0)$ into the point at infinity and taking the x-axis into itself. The result is shown in Figure 2. The two solid circles are the transforms of the circles of curvature at $(\pm a, 0)$ and, as the curve is directed, are the two circles of minimum curvature. Let them be denoted by C_1 and C_2 as shown. Consider the curve C obtained by tracing arc ABD, then tracing C_1 n times, then arc DEA, and finally C_2 n times, the directions being as indicated. As in Section **9**, consider this curve on the logarithmic Riemann surface with branch point Q, as shown. Since the total rotation of the vector QP as P traces C is clearly zero, C is a closed curve on this surface. Moreover, it is easily verified that on this surface it is simple, for the various double points in the plane correspond to points on different sheets of the surface. C is therefore a simple closed curve with exactly four vertices (since that is true of the ellipse) lying in a simply connected region of a surface of constant (zero) curvature. The simply connected region is that part of the surface covering an annular region of the plane about Q.

Consider the dotted circle K in the figure. Since it goes around the branch point, it is an open arc and has points on all sheets of the Riemann surface. The same is true of C_1, and C_1 and K have two intersections on each sheet of the surface. Since curve C contains the part of C_1 on n sheets, C meets K $2n$ times. Moreover, the order of the points is the same on C and K. This proves that Theorem 11. 2 is false if the requirement that the geodesic circle be complete is removed, for if the theorem were true it would say that C has $2n$ geodesic vertices, and this is false when n exceeds 2. This answers question (a).

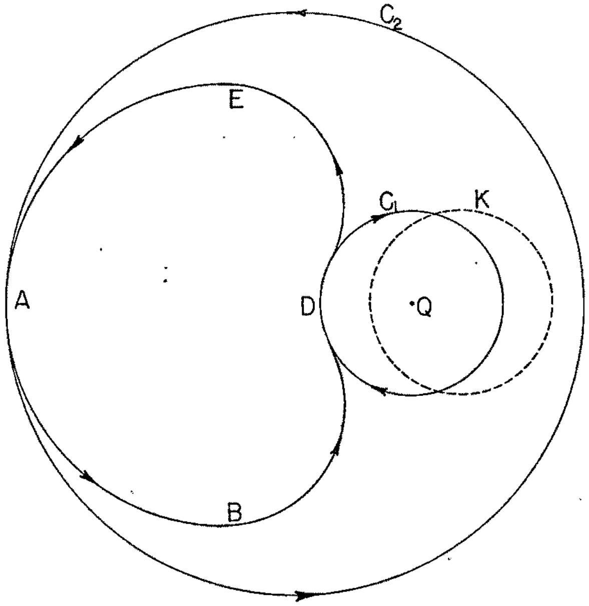


FIGURE 2.

To answer question (b) consider the trisectrix whose polar equation is $r = 1 + 2 \operatorname{Cos} \theta$. The smallest circle C which contains this curve has two points of contact with it which divides the trisectrix into two arcs. Let AB be the one of these arcs containing the double point, and let it be directed so that the loop is positively traced. The only vertex on this arc is therefore a maximum. As in the last example, let this be considered on a logarithmic

Riemann surface with branch point at a point O interior to the loop, and therefore also to C. On this surface C is an open arc with points on all sheets. Moreover, on this surface arc AB is simple since the double point corresponds to points on different sheets of the surface. On the surface, therefore, AB is a simple arc tangent to a geodesic circle in the same sense at A and B and lying to the left of this directed circle, yet it has no minimum of the geodesic curvature as there would have to be if Lemma 5. 2 were true in this case. Indeed the example is a counterexample even for the simpler Lemma 5. 1. Question (b) above must therefore also be answered in the negative.

UNIVERSITY OF MARYLAND.

BIBLIOGRAPHY.

1. W. Blaschke, *Kreis and Kugel*, Leipzig, 1916, p. 161.

2. ———, *Vorlesungen über Differentialgeometrie I*, Berlin, 1930.

3. G. Bol, "Ein Satz über Eilinien," *Abh. Math. Sem. Hansischen Univ.*, vol. 13 (1940), pp. 319-320.

4. D. Fog, "Über den Vierscheitelsatz und seine Verallgemeinerungen," *Sitz. der Berlin Akademie der Wissenschaft* (1933), pp. 251-254.

5. W. C. Graustein, "Extensions of the four-vertex theorem," *Transactions of the American Mathematical Society*, vol. 41 (1937), pp. 9-23.

6. S. B. Jackson, "Vertices of plane curves," *Bulletin of the American Mathematical Society*, vol. 50 (1944), pp. 564-578.

7. ———, "The four-vertex theorem for surfaces of constant curvature," *American Journal of Mathematics*, vol. 67 (1945), pp. 563-582.

8. H. Mohrmann, "Die Minimalzahl der stationären Ebenen einer räumlichen Ovals," *Sitz. Math.-Nat. Abt. Bayer. Akad. Wiss.* (1917), pp. 1-4.

9. P. Scherk, "The four-vertex theorem," *Proceedings of the First Canadian Mathematical Congress*, Montreal 1945, Toronto, 1946, pp. 97-102.

10. H. Hopf, "Über die Drehung der Tangenten und Sehnen ebener Kurven," *Compositio Math.*, vol. 2 (1935), pp. 50-62.

THE GENERAL TERM OF THE GENERALIZED SCHLÖMILCH SERIES.*

By J. Ernest Wilkins, Jr.

1. Introduction. By a generalized Schlömilch series is meant a series of the form

$$\tfrac{1}{2}a_0/\Gamma(\nu+1)+\sum_{m=1}^{\infty}(\tfrac{1}{2}mx)^{-\nu}\{a_mJ_\nu(mx)+b_mH_\nu(mx)\},$$

in which $J_\nu(u)$ is the Bessel function of first kind and order ν,

$$J_\nu(u)=\sum_{n=0}^{\infty}(-1)^n(\tfrac{1}{2}u)^{\nu+2n}/n!\Gamma(\nu+n+1),$$

and $H_\nu(u)$ is the Struve function of order ν,

$$H_\nu(u)=\sum_{n=0}^{\infty}(-1)^n(\tfrac{1}{2}u)^{\nu+2n+1}/\Gamma(n+3/2)\Gamma(\nu+n+3/2).$$

Watson [2; 645] has shown that if the general term of the above generalized Schlömilch series converges to zero for all values of x in any interval, then $a_m=o(m^{\nu+\frac{1}{2}})$, $b_m=o(m^{\nu+\frac{1}{2}})$, provided that ν is a real number less than $\frac{1}{2}$.

By analogy with the Cantor lemma [1; 84] we would expect to be able to prove this assertion even if the interval were replaced by an arbitrary set E of positive measure. This is our first result (Theorem 1). By making use of the explicit formulas for $J^{\frac{1}{2}}(u)$ and $H^{\frac{1}{2}}(u)$ it is next seen (Theorem 2) that this result is still true when $\nu=\frac{1}{2}$. Even more is true when $\nu>\frac{1}{2}$; we shall prove that $a_m=o(m^{\nu+\frac{1}{2}})$, $b_m=o(m)$, in this case (Theorem 3).

2. The case when $\nu<\frac{1}{2}$. In this section we shall prove the following theorem.

Theorem 1. *Suppose that* $-\infty<\nu<\frac{1}{2}$ *and that*

$$(\tfrac{1}{2}mx)^{-\nu}\{a_mJ_\nu(mx)+b_mH_\nu(mx)\}=o(1)$$

for all x in a set E of positive measure. Then $a_m=o(m^{\nu+\frac{1}{2}})$, $b_m=o(m^{\nu+\frac{1}{2}})$.

Since $u^{-\nu}J_\nu(u)$ and $u^{-\nu}H_\nu(u)$ are respectively even and odd functions of u, we can suppose that in the proof of Theorem 1 (and the subsequent theorem also) E is contained in an interval (α,β) such that $0<\alpha<\beta$.

If Theorem 1 were false there would be an infinite subset M of the set of positive integers and a positive quantity η such that $|a_m|^2+|b_m|^2>\eta^2m^{2\nu+1}$

---

* Received August 24, 1948.

if m is in M. Since $0 < \alpha \leqq x \leqq \beta$ if x is in E, it follows that if $f_m(x) = (\pi m x)^{\frac{1}{2}}\{A_m J_\nu(mx) + B_m H_\nu(mx)\}$, where $A_m = a_m/(|a_m|^2 + |b_m|^2)^{\frac{1}{2}}$, $B_m = b_m/(|a_m|^2 + |b_m|^2)^{\frac{1}{2}}$, then $f_m(x) = o(1)$ for all x in E as m approaches ∞ in M. It is known [2;199] that

$$\begin{aligned} J_\nu(u) &= (2/\pi u)^{\frac{1}{2}}[\cos(u - \tfrac{1}{2}\nu\pi - \tfrac{1}{4}\pi) + O(u^{-1})], \\ Y_\nu(u) &= (2/\pi u)^{\frac{1}{2}}[\sin(u - \tfrac{1}{2}\nu\pi - \tfrac{1}{4}\pi) + O(u^{-1})], \end{aligned} \tag{2.1}$$

as u approaches ∞, and that [2;333]

$$H_\nu(u) = Y_\nu(u) + (\tfrac{1}{2}u)^{\nu-1}/\pi^{\frac{1}{2}}\Gamma(\nu + \tfrac{1}{2}) + O(u^{\nu-3}) \tag{2.2}$$

as u approaches ∞. It follows from these relations and the fact that E is bounded away from zero that $J_\nu(mx) = O(m^{-\frac{1}{2}})$ and $H_\nu(mx) = O(m^{-\frac{1}{2}})$ when $\nu < \frac{1}{2}$, and the constant implied by the symbols O can be chosen independent of x. It follows that $f_m(x)$ is bounded, and hence that

$$\int_E |f_m(x)|^2 dx = o(1) \tag{2.3}$$

as m approaches ∞ in M. If we use equations (2.1) and (2.2) we find that

$$\begin{aligned} |f_m(x)|^2 = 1 &+ (|A_m|^2 - |B_m|^2)\sin(2mx - \nu\pi) \\ &- 2Re(A_m B_m^*)\cos(2mx - \nu\pi) + o(1), \end{aligned} \tag{2.4}$$

in which B_m^* is the complex conjugate of B_m, $Re(u)$ is the real part of u, and the term $o(1)$ is uniform in x on E. Since $|A_m| \leqq 1$ and $|B_m| \leqq 1$, it follows upon substituting equation (2.4) into equation (2.3) and using the Riemann-Lebesgue lemma that the measure of E is zero, and this contradicts the hypothesis of the theorem. This contradiction completes the proof of Theorem 1.

3. The case when $\nu = \frac{1}{2}$.

In this section we shall prove the following theorem.

THEOREM 2. *The conclusion of Theorem* 1 *follows from its hypothesis if* $\nu = \frac{1}{2}$.

It is known [2;54, 333] that

$$J_{\frac{1}{2}}(mx) = (2/\pi mx)^{\frac{1}{2}}\sin mx, \qquad H_{\frac{1}{2}}(mx) = (2/\pi mx)^{\frac{1}{2}}(1 - \cos mx).$$

The hypothesis of Theorem 1 thus reduces to the assumption that

$$\begin{aligned} &(mx)^{-1}\{a_m \sin mx + b_m(1 - \cos mx)\} \\ &\qquad = 2(mx)^{-1}\sin\tfrac{1}{2}mx\{a_m\cos\tfrac{1}{2}mx + b_m\sin\tfrac{1}{2}mx\} = o(1) \end{aligned}$$

for all x in E.

If Theorem 2 were false there would be an infinite subset M of the set of positive integers and a positive quantity η such that $|a_m|^2 + |b_m|^2 > \eta^2 m^2$ whenever m is in M. If E is bounded and bounded away from zero as in the previous section, and if A_m and B_m are defined by equations (2.0), it follows that $f_m(x) = 2(A_m \cos \frac{1}{2}mx + B_m \sin \frac{1}{2}mx) \sin \frac{1}{2}mx = o(1)$ for all x in E as m approaches ∞ in M. Since $f_m(x)$ is plainly bounded, we have (2.3) as m approaches ∞ in M. It is easy to see that

$$|f_m(x)|^2 = \tfrac{1}{2}(|A_m|^2 + 3|B_m|^2) - 2|B_m|^2 \cos mx + 2Re(A_m B_m^*) \sin mx$$
$$- Re(A_m B_m^*) \sin 2mx - \tfrac{1}{2}(|A_m|^2 - |B_m|^2) \cos 2mx.$$

It then follows from equation (2.3) and the Riemann-Lebesgue lemma that $|A_m|^2 + 3|B_m|^2 = o(1)$ as m approaches ∞ in M, since E has positive measure. This is impossible, however, since $|A_m|^2 + 3|B_m|^2 \geqq |A_m|^2 + |B_m|^2 = 1$. This contradiction completes the proof of Theorem 2.

4. The case when $\nu > \frac{1}{2}$. In this section we shall prove the following theorem.

THEOREM 3. *If* $\nu > \frac{1}{2}$ *and the hypothesis of Theorem* 1 *holds, then* $a_m = o(m^{\nu+\frac{1}{2}})$, $b_m = o(m)$.

If Theorem 3 were false there would be an infinite subset M of the set of positive integers and a positive quantity η such that $c_m^2 = |a_m|^2 m^{-2\nu-1} + |b_m|^2 m^{-2} > \eta^2$ if m is in M. Let us define $A_m = a_m / m^{\nu+\frac{1}{2}} c_m$, $B_m = b_m / m^{\nu+\frac{1}{2}} c_m$. Then

$$(4.0) \qquad |A_m| \leqq 1, \qquad |B_m| \leqq m^{\frac{1}{2}-\nu}.$$

If we suppose as in the preceding sections that E is bounded and bounded away from zero it then follows that for every value of μ,

$$f_{m\mu}(x) = x^{(1+\mu)/2}(m\pi)^{\frac{1}{2}}\{A_m J_\nu(mx) + B_m H_\nu(mx)\} = o(1)$$

as m approaches ∞ in M. To see that $f_{m\mu}(x)$ is bounded we use the inequalities (4.0) and notice from equations (2.1) and (2.2) that $J_\nu(mx) = O(m^{-\frac{1}{2}})$ and $H_\nu(mx) = O(m^{\nu-1})$ uniformly in x on E when $\nu > \frac{1}{2}$. It follows that

$$\int_E |f_{m\mu}(x)|^2 dx = o(1)$$

as m approaches ∞ in M. If we define $\phi_\nu(mx)$ so that $H_\nu(mx) = Y_\nu(mx) + \phi_\nu(mx)$, then it follows from equation (2.2) that

$$(4.1) \qquad \phi_\nu(mx) - (\tfrac{1}{2}mx)^{\nu-1}\pi^{-\frac{1}{2}}/\Gamma(\nu + \tfrac{1}{2}) = O(m^{\nu-3}),$$

and that $|f_{m\mu}(x)|^2 = \pi m x^{1+\mu}[|A_m J_\nu(mx) + B_m Y_\nu(mx)|^2 - |B_m|^2 \phi_\nu^2(mx) + 2Re\{B_m^* \phi_\nu(mx)[A_m J_\nu(mx) + B_m H_\nu(mx)]\}]$. If we use equations (2.2) and (4.1) we find that

$$|f_{m\mu}(x)|^2 = x^\mu[|A_m|^2 + |B_m|^2 + (|A_m|^2 - |B_m|^2)\sin(2mx - \nu\pi)$$
$$- |B_m|^2\{(mx)^{2\nu-1}u_\nu^{-1} + O(m^{2\nu-3})\} - 2Re(A_mB_m^*)\cos(2mx - \nu\pi)$$
$$+ 2Re\{B_m^*\phi_\nu(mx)[A_mJ_\nu(mx) + B_mH_\nu(mx)]\} + o(1)],$$

in which $u_\nu = 2^{2\nu-2}\Gamma^2(\nu + \frac{1}{2})$. Since (4.0) holds, it follows from the Riemann-Lebesgue lemma that the integral of $x^\mu[(|A_m|^2 - |B_m|^2)\sin(2mx - \nu\pi) - 2Re(A_mB_m^*)\cos(2mx - \nu\pi)]$ over E converges to zero as m approaches ∞ in M. Since $B_m^*\phi_\nu(mx) = O(m^{-\frac{1}{2}})$ uniformly on E, it is true that $x^\mu B_m^*\phi_\nu(mx)[A_mJ_\nu(mx) + B_mH_\nu(mx)] = O(m^{-1})$ uniformly on E as m approaches ∞ in M, and hence that the integral of this expression converges to zero. Similarly, the term $|B_m|^2O(m^{2\nu-3}) = O(m^{-2})$ has an integral over E which converges to zero as m approaches ∞ in M. We thus conclude that

$$\int_E x^\mu(|A_m|^2 + |B_m|^2 - |B_m|^2m^{2\nu-1}x^{2\nu-1}u_\nu)\,dx = o(1)$$

as m approaches ∞ in M. Since $|A_m|^2 + |B_m|^2 \leqq 1$ and $|B_m|^2m^{2\nu-1} \leqq 1$, we may replace M by an infinite subset for which there exist quantities A and B such that

$$\lim(|A_m|^2 + |B_m|^2) = A, \qquad \lim |B_m|^2m^{2\nu-1} = B.$$

Evidently, $0 \leqq A \leqq 1$, $0 \leqq B \leqq 1$, and

$$\int_E x^\mu(A - Bu_\nu x^{2\nu-1})\,dx = 0.$$

Since μ is arbitrary we infer that $A - Bu_\nu x^{2\nu-1} = 0$ almost everywhere on E and hence that $A = B = 0$ since E has positive measure. It follows, however, from the definitions of A_m and B_m that

$$|A_m|^2 + |B_m|^2(1 + m^{2\nu-1}) = 1 + |b_m|^2/(|a_m|^2 + |b_m|^2m^{2\nu-1}) \geqq 1,$$

whence $A + B \geqq 1$ and so it cannot happen that $A = B = 0$. This contradiction completes the proof of the theorem.

BIBLIOGRAPHY.

1. G. H. Hardy and W. W. Rogosinski, *Fourier series*, Cambridge Tracts in Mathematics and Mathematical Physics, No. 38, Cambridge, 1944.
2. G. N. Watson, *A treatise on the theory of Bessel functions*, Cambridge, 1945.

ON THE EXTENSION OF THE PARTIAL ORDER OF GROUPS.*

By LADISLAS FUCHS.

1. In his paper "Sur l'extension de l'ordre partiel"[1] E. Szpilrajn has proved that every partial order defined on a set has a linear extension. This general result does not necessarily hold for partially ordered groups, since it may obviously happen that the extended order does not satisfy the group axioms. The principal purpose of the present paper is the demonstration of the same theorem on abelian groups. The theorem will be proved to hold only for groups on which an additional condition is satisfied, a condition which requires that only positive elements have positive natural multiples.

2. We recall that an abelian group G, written additively, is said to be a *partially ordered group*[2] if a relation $>$ is defined between some pairs of its elements such that the following postulates hold:

(i) *any two of the three relations* $a > b$, $a = b$, $a < b$ *are contradictory;*

(ii) *transitivity:* $a > b$ *and* $b > c$ *imply* $a > c$;

(iii) *homogeneity:* $a > b$ *implies* $a + c > b + c$ *for every* c *in* G.

By the laws (ii) and (iii) the relations $a > b$, $c > d$ may be added to get $a + c > b + d$.

If in addition G satisfies the condition:

(iv) $na = a + a + \cdots + a \geqq 0$ *for some positive integer* n *implies* $a \geqq 0$,

we say the partial order is *normal.*

If G is such that any two elements a, b are *comparable* in the sense that one of the possibilities $a > b$, $a = b$, $a < b$ does hold, we say G is *linearly ordered.* A linearly ordered group always satisfies condition (iv), for if, under the hypothesis $na \geqq 0$, $a \geqq 0$ did not hold, we should then have by linear order $a < 0$ implying $na < 0$ for every positive integer n.

We now prove that a group on which a normal partial order is defined has, with the exception of 0, only elements of infinite order. In fact, the normality states that if $na = 0$, or, what is the same, if $na \geqq 0$ and $-na \geqq 0$, then $a \geqq 0$ as well as $-a \geqq 0$, that is, $a = 0$.

---

* Received February 3, 1949.

[1] *Fundamenta Mathematicae*, vol. 16 (1930), pp. 386-389.

[2] Our definition is stated in a form which is most convenient for our purpose; cf. C. J. Everett and S. Ulam, "On ordered groups," *Transactions of the American Mathematical Society*, vol. 57 (1945), pp. 208-216.

Suppose that two partial orders P and R are defined on the same group and a relation $a > b$ in P implies $a > b$ in R; then R will be called an *extension* of P. An extension which defines a linear order on G will be called a *linear extension.*

3. We shall now prove the following lemma.

LEMMA. *If P is a normal partial order on the group G and x and y are any two elements non-comparable in P, then there exists an extension R of P such that $x > y$ in R. Moreover if such an extension may be carried out for any two non-comparable elements, then P is necessarily normal.*[3]

For the proof assume that P is a normal partial order on G and the elements x and y are not comparable in P. Let us define a relation R as follows.

We put $a > b$ in R if and only if $a \neq b$ and there are two non-negative integers p, q, not both zero, such that

$$p(a-b) \geqq q(x-y) \text{ in } P. \tag{1}$$

What we have to show is that by this definition R is a partial order and an extension of P, further $x > y$ in R.

First of all, we note that p is never zero, for otherwise we should have $0 \geqq q(x-y)$ in P for a certain positive integer q, whence by (iv) we have $y \geqq x$ in P against hypothesis.

α) We begin with verifying condition (i) for R. It is clearly enough to show that $a > b$ in R and $b > a$ in R are contradictory. For assume $p(a-b) \geqq q(x-y)$ in P, as well as $r(b-a) \geqq s(x-y)$ in P, for some non-negative integers p, q, r, s. By adding r times the first, p times the second inequality, one obtains $pr(a-b) + pr(b-a) \geqq (qr+ps)(x-y)$ in P, that is to say, $0 \geqq (qr+ps)(x-y)$ in P. If $qr+ps$ does not vanish, by normality we are led to $y \geqq x$ in P, a contradiction. On the other hand, if $qr+ps$ is zero, i. e., both q and s vanish, then the inequalities $p(a-b) \geqq 0$ in P and $r(b-a) \geqq 0$ in P imply by normality $a \geqq b$ in P and $b \geqq a$ in P, i. e., $a = b$ which is absurd.

β) We proceed now to the proof of the transitivity of R. Assume that $a > b$ in R and $b > c$ in R, that is, for some non-negative integers p, q, r, s, $p(a-b) \geqq q(x-y)$ in P and $r(b-c) \geqq s(x-y)$ in P. By adding as in α) one gets $pr(a-b) + pr(b-c) = pr(a-c) \geqq (qr+ps)(x-y)$ in P. Here pr is not zero and $a = c$ is by α) impossible, so that, by definition, $a > c$ in R, which establishes the transitivity of R.

---

[3] Our proof is in some respect similar to Szpilrajn's proof.

γ) In order to prove the homogeneity for R, take into account that (1) includes only the difference $a - b$ which is equal to $(a + c) - (b + c)$ for every c in G and $a + c = b + c$ is impossible if a and b are different.

δ) R is an extension of P, for if $a > b$ in P, then $a - b > 0$ in P, consequently, for $p = 1$, $q = 0$, condition (1) is satisfied, hence $a > b$ in R.

ϵ) One sees at once that $x > y$ in R. In fact, for $x = a$, $y = b$ and $p = q = 1$, the relation (1) takes the form $x - y \geqq x - y$ in P.

ξ) To complete the proof of the lemma, let us now assume that there is an element g in G such that $ng \geqq 0$ in P without $g \geqq 0$ in P. Then g and 0 are not comparable in P and we may infer, from the above discussions, that there exists an extension of P in which $g < 0$. This is however absurd, since this would imply $ng < 0$ in R, contrary to the hypothesis $ng \geqq 0$ in P.

4. We may prove even the normality for the partial order R defined in **3**. Indeed, supposing $na \geqq 0$ in R for some positive integer n, i. e., $p(na) = (pn)a \geqq q(x - y)$ in P, we are led at once to the result $a \geqq 0$ in R.

What has been proved shows that the extended partial order R may again be extended to another partial order S, in which two prescribed elements non-comparable in R become comparable, etc.

If, in general, $P_1, P_2, \cdots, P_\tau, \cdots$ is a well-ordered chain of partial orders such that each of them is some extension of the preceding ones, then the *union* of the chain may be defined to be a partial order P such that $a > b$ in P if and only if $a > b$ in P_τ holds for one and hence for all subsequent subscripts τ. There is no difficulty in establishing that P is normal if all P_τ are normal.

Hence, as a simple consequence of Zorn's lemma[4] we get that in the set of all normal partial orders defined on the group G which are extensions of P, there are maximal orders M, that is, orders which have no proper extension. By our Lemma this can happen only in case any two elements are comparable in M, that is to say, M is a linear order. Thus we have, immediately, the result:

THEOREM 1. *For every normal partial order P defined on G and every two elements x, y non-comparable in P, there exists a linear extension L_{xy} with the property that $x > y$ in L_{xy}.*

---

[4] M. Zorn, "A remark on method in transfinite algebra," *Bulletin of the American Mathematical Society*, vol. 41 (1935), pp. 667-670.

5. Let $\mathfrak{S} = \{P_1, P_2, \cdots, P_\tau, \cdots\}$ be any set of partial orders, each defined on the same group G. We define a new partial order P on G as follows. For any two elements a, b we put $a > b$ in P if and only if $a > b$ in every partial order P_τ in the set $\mathfrak{S}$. It is readily seen that P is again a partial order, moreover, P is normal if all partial orders of $\mathfrak{S}$ are normal. The partial order P is said to be the *product* of the P_τ or to be *realized* by the set $\mathfrak{S}$ of partial orders, written $P = \Pi P_\tau$.

THEOREM 2. *A partial order P defined on a group G may be realized by a certain set of linear orders if and only if P is normal.*

The necessity is obvious, since a linear order, and hence every product of linear orders, is normal. On the other hand, if P is not itself linear, then take to any pair of elements x, y non-comparable in P the corresponding linear extensions L_{xy} and L_{yx} described in Theorem 1. It is easily seen that these linear orders realize P.

If by the dimension[5] of a partial order P we mean the least cardinal number $\mathfrak{w}$ such that P may be realized by $\mathfrak{w}$ linear orders, then we can reformulate Theorem 2:

THEOREM 3. *A partial order P defined on a group has a dimension if and only if P is normal.*

This theorem states, for example, that each commutative group which is lattice-ordered in the sense of G. Birkhoff,[6] has a well-defined dimension.

6. When we start with an abelian group on which the relation $>$ is defined for no pair of elements, then, by applying Theorem 2 we come to a theorem due to F. Levi.[7]

THEOREM 4. *In a commutative group a linear order may be defined, if all of its elements, except* 0, *are of infinite order.*

Indeed, a group in which no partial order is defined is normal if and only if it has no element of finite order other than 0.

BUDAPEST, HUNGARY.

---

[5] This definition is due to Ben Dushnik and E. W. Miller, "Partially ordered sets," *American Journal of Mathematics*, vol. 63 (1941), pp. 600-610.

[6] G. Birkhoff, "Lattice-ordered groups," *Annals of Mathematics* (2), vol. 43 (1942), pp. 298-331. Lemma 3 of §9 states that a lattice-ordered abelian group is always normal.

[7] F. Levi, "Arithmetische Gesetze im Gebiete diskreter Gruppen," *Rendiconti Palermo*, vol. 35 (1913), pp. 225-236.

ON THE CONSTRUCTION OF PARTIALLY ORDERED SYSTEMS WITH A GIVEN GROUP OF AUTOMORPHISMS.*

By Robert Frucht.

In a recent paper [1] G. Birkhoff showed that when an abstract group of g elements is given (g being a finite number or an infinite cardinal number), there exists always a partially ordered system with $g^2 + g$ elements whose group of automorphisms is simply isomorphic to the given group. I shall prove the following result for the case of a finite g: *There can be found a partially ordered system with only* $(n + 2)g$ *elements whose group of automorphisms is simply isomorphic to a given abstract group of finite order* g, *when this group can be generated by* n *of its elements*; and if $n > 2$, still fewer elements will be needed.

Proof. Let a_1 be the identity of the given group $\mathfrak{G}$, and $a_2, a_3, \cdots, a_{n+1}$ the n generating elements; $a_{n+2}, a_{n+3}, \cdots, a_g$ be the other elements of $\mathfrak{G}$. We proceed now to construct a partially ordered system with the following $(n+2)g$ elements: the g maximal elements (a_1), (a_2), $(a_3), \cdots, (a_g)$ corresponding to the g elements of the group $\mathfrak{G}$, and the $(n+1)g$ other elements (a_ρ, a_σ) which correspond to ordered pairs of elements of $\mathfrak{G}$; here the "first component" a_ρ (where $\rho = 1, 2, \cdots, g$) corresponds to any element of $\mathfrak{G}$, but the "second component" a_σ (where $\sigma = 1, 2, \cdots, n+1$) corresponds either to the unit a_1 or to any *generating* element $a_2, a_3, \cdots, a_{n+1}$ of $\mathfrak{G}$. (This limitation of the second component to the generating elements of $\mathfrak{G}$ just represents our modification of Birkhoff's original method by which the number of elements is reduced from $(g+1)g$ to $(n+2)g$.) Between these $(n+2)g$ elements the following covering relations are defined: [2]

---

* Received May 11, 1949.

[1] Garrett Birkhoff, "Sobre los grupos de automorfismos," *Revista de la Unión Matemática Argentina*, vol. 11 (1946), pp. 155-157; see also: R. Frucht, "Sobre la construcción de sistemas parcialmente ordenados con grupo de automorfismos dado," *Revista de la Unión Matemática Argentina*, vol. 13 (1948), pp. 12-18.

[2] It is not necessary to postulate (b) also for $\tau = 1$, as a_1 is the unit of $\mathfrak{G}$, and therefore $a_1 a_\sigma = a_\sigma$, so that (b) for $\tau = 1$ reads: $(a_\sigma) > (a_\sigma, a_1)$; but this relation is already contained in (a).

$$\begin{cases} \text{(a)} \quad (a_\rho) > (a_\rho, a_1) > (a_\rho, a_2) > \cdots > (a_\rho, a_{n+1}) & \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (\text{for } \rho = 1, 2, \cdots, g); & \\ \text{(b)} \quad (a_\sigma) > (a_\tau a_\sigma, a_\tau) \qquad (\text{for } \sigma = 1, 2, \cdots, g; \tau = 2, 3, \cdots, n+1). & \end{cases}$$

It is obvious that the system thus defined is a finite partially ordered system (in the usual[3] terminology), in which no element is isolated. We are now going to prove that this system has a group of automorphisms simply isomorphic to $\mathfrak{G}$.

Let λ be any number from $1, 2, \cdots, g$, and consider the following mapping of the partially ordered system defined by (a) and (b) into itself:

$$\text{(c)} \qquad \Phi_\lambda \begin{cases} (a_\rho) \rightarrow (a_\rho a_\lambda) \\ (a_\sigma, a_\tau) \rightarrow (a_\sigma a_\lambda, a_\tau). \end{cases}$$

This mapping Φ_λ is clearly order-preserving; e. g. Φ_λ carries (a_σ) into $(a_\sigma a_\lambda)$ and $(a_\tau a_\sigma, a_\tau)$ into $(a_\tau a_\sigma a_\lambda, a_\tau)$, whence the relations (b) are changed into the (likewise correct) relations: $(a_\sigma a_\lambda) > (a_\tau a_\sigma a_\lambda, a_\tau)$.

Now let λ run through $1, 2, \cdots, g$; then the formulae (c) will give g distinct automorphisms $\Phi_1, \Phi_2, \cdots, \Phi_g$ of the partially ordered system, and from the first line of (c) it is obvious that these mappings constitute a group simply isomorphic to $\mathfrak{G}$.

It remains to be shown that there are no other automorphisms of the system besides those given by (c) for $\lambda = 1, 2, \cdots, g$. Let Φ be any order-preserving mapping of the system into itself; we are to prove that Φ coincides with one of the g mappings Φ_λ. According to (a) and (b), the g elements $(a_1), (a_2), \cdots, (a_g)$ are the only maximal elements of the system; hence Φ can merely permute them among themselves. Let (a_μ) be that maximal element into which (a_1) is carried by Φ; and let Φ_μ^{-1} be the automorphism inverse to Φ_μ. Since Φ_μ changes (a_1) into $(a_1 a_\mu) = (a_\mu)$, Φ_μ^{-1} carries (a_μ) into (a_1); hence the product $\Phi\Phi_\mu^{-1}$ (i. e. the mapping Φ followed by the mapping Φ_μ^{-1}) leaves the element (a_1) fixed. We will show that this mapping leaves unchanged all the elements of the partially ordered system.

Let us begin with the maximal chain (of length $n+1$) that starts with (a_1): $(a_1) > (a_1, a_1) > (a_1, a_2) > (a_1, a_3) > \cdots > (a_1, a_{n+1})$. According to (b), (a_1) covers also the elements $(a_2, a_2), (a_3, a_3), \cdots, (a_{n+1}, a_{n+1})$, but since the second component of each of these elements is different from a_1, there is no other maximal chain beginning with (a_1); and (a_1, a_1), (a_1, a_2),

---

[3] Garrett Birkhoff, *Lattice Theory*, first ed., p. 5.

$\cdots, (a_1, a_{n+1})$ are left fixed by $\Phi\Phi_\mu^{-1}$, as they constitute the only maximal chain starting with (a_1).

Consider now the elements (a_τ, a_τ), with $\tau = 2, 3, \cdots, n+1$, which are covered by (a_1); they could only be interchanged by $\Phi\Phi_\mu^{-1}$; but the largest chain starting with (a_τ, a_τ) is

$$(a_\tau, a_\tau) > (a_\tau, a_{\tau+1}) > (a_\tau, a_{\tau+2}) > \cdots > (a_\tau, a_{n+1})$$

and hence of length $n+1-\tau$, and since this length is different for any two values of τ, no such interchange of the elements (a_τ, a_τ) is possible. It follows that also the elements $(a_2, a_2), (a_3, a_3), \cdots, (a_{n+1}, a_{n+1})$ are left fixed by $\Phi\Phi_\mu^{-1}$.

But since any one of these elements belongs only to one maximal chain (of length $n+1$), viz. (a_τ, a_τ) to the maximal chain:

$$(a_\tau) > (a_\tau, a_1) > (a_\tau, a_2) > \cdots > (a_\tau, a_\tau) > \cdots > (a_\tau, a_{n+1}),$$

also all the other elements of all these maximal chains are left fixed by $\Phi\Phi_\mu^{-1}$.

Thus we have already recognized as fixed the elements (a_σ) and (a_σ, a_τ) for $\sigma = 1, 2, \cdots, n+1$; $\tau = 1, 2, \cdots, n+1$. It remains to be proved that the same is true also for $\sigma = n+2, n+3, \cdots, g$. This may be accomplished by the following reasoning:

According to (b) any element (a_σ)—already recognized as fixed if $\sigma = 1, 2, \cdots, n+1$—covers the elements

$$(a_2 a_\sigma, a_2), (a_3 a_\sigma, a_3), \cdots, (a_{n+1} a_\sigma, a_{n+1}),$$

which must be left unchanged by $\Phi\Phi_\mu^{-1}$ too (due to the different lengths of the longest chain starting with each of them); and with them also the other elements of the maximal chains containing them must remain fixed, i. e. all the elements $(a_\rho a_\sigma)$ and $(a_\rho a_\sigma, a_\tau)$, where $a_\rho a_\sigma$ is a product of any two *generating* elements of the group $\mathfrak{G}$.

Repeating this reasoning for the elements covered by $(a_\rho a_\sigma)$ and their maximal chains, we are led to the conclusion that also the elements $(a_\pi a_\rho a_\sigma)$ and $(a_\pi a_\rho a_\sigma, a_\tau)$ are left unchanged by $\Phi\Phi_\mu^{-1}$, where $a_\pi a_\rho a_\sigma$ is any product of 3 generating elements of $\mathfrak{G}$; and continuing the proof in the same way (or shortening it by complete induction as to the number of generating elements appearing as factors), we will easily recognize as fixed *all* the elements (a_σ) and (a_σ, a_τ), since any a_σ is the product of a finite number of generating elements of $\mathfrak{G}$.

The proof is now readily completed as follows: We have already shown that $\Phi\Phi_\mu^{-1}$ leaves all the elements of our system unchanged; hence $\Phi\Phi_\mu^{-1} = \Phi_1$ (identity), and $\Phi = \Phi_1\Phi_\mu = \Phi_\mu$. Thus we have proved that any automorphism of the system coincides with one of the mappings Φ_μ given by (c), and we already know that these g mappings constitute a group simply isomorphic to the given abstract group $\mathfrak{G}$.

Finally we shall prove that if $n > 2$ *the system having the covering relations* (a) *and* (b) *defined above may be modified so that less than* $(n+2)g$ *elements are needed.* To prove this it will suffice to notice that if $n > 2$ it is possible to drop all the elements (a_ρ, a_σ) where σ is greater than a number m depending on n and defined as the least integer satisfying the inequality: $m \geqq \frac{1}{2}[1 + (8n+1)^{\frac{1}{2}}]$. One must only limit the covering relations (a) to the retained elements, and replace those of the relations (b) which refer to dropped elements, by others of the type:

$$(b') \qquad (a_\rho, a_\kappa) > (a_\sigma a_\rho, a_\lambda),$$

where σ runs through $m+1, m+2, \cdots, n+1$, and where for any two values of σ two distinct combinations $\kappa < \lambda$ (from the numbers $1, 2, \cdots, m$) must be chosen. In order to avoid the formation of new maximum chains it will be convenient to choose always $\kappa \leqq \lambda - 2$. As there are $\frac{1}{2}(m-1)(m-2)$ combinations satisfying this condition, and $n - m + 1$ dropped suffixes, m can be chosen as the least positive integer satisfying the inequality:

$$\tfrac{1}{2}(m-1)(m-2) \geqq n - m + 1,$$

whence the value of m given above follows.

The proof that also the modified system of $(m+1)g$ elements has a group of automorphisms simply isomorphic to the given group $\mathfrak{G}$ may easily be supplied by the reader along the same lines of the proof given above for the system defined by (a) and (b).

Addendum.

Of course it will be convenient to take n always as small as possible, e. g. $n = 1$ for cyclic groups, $n = 2$ for dihedral groups, etc. In any group of finite order g however the number n of generators can always be chosen so small that $n \leqq \log g/\log 2$; hence we have the following

COROLLARY. *To obtain a partially ordered system whose group is simply isomorphic to a given group of finite order g, at most $[(2+\log g/\log 2)g]$ elements are needed.*[4]

Here $[\ldots]$ stands as usual for "greatest integer $\leqq$."

It may be noted that when the factorization of g into prime powers is known: $g = p_1^{\alpha_1}p_2^{\alpha_2}p_3^{\alpha_3}\cdots p_k^{\alpha_k}$, it can be shown that

$$n \leqq \alpha_1 + \alpha_2 + \alpha_3 + \cdots + \alpha_k,$$

whence another corollary similar to the foregoing can be concluded.

TECHNICAL UNIVERSITY F. SANTA MARIA,
VALPARAISO, CHILE, SOUTH AMERICA.

---

[4] Combining the two inequalities: $m < \frac{1}{2}\{3 + (8n+1)^{\frac{1}{2}}\}$ and $n \leqq \log g/\log 2$, it follows that in this corollary $[(2+\log g/\log 2)g]$ can be replaced by $[(5+\gamma)g/2]$, where $\gamma = (1 + 8\log g/\log 2)^{\frac{1}{2}}$.

ON THE BEHAVIOUR OF FOURIER TRANSFORMS AT INFINITY AND ON QUASI-ANALYTIC CLASSES OF FUNCTIONS.*[1]

By I. I. Hirschman, Jr.

1. Introduction. Let $f(x) \in L_2(-\infty, \infty)$ and $\phi(t) \in L_2(-\infty, \infty)$ be corresponding Fourier transforms,

$$(1) \qquad \begin{aligned} f(x) &= (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} e^{ixt}\phi(t)\,dt \qquad (\mathrm{M}_2) \\ \phi(t) &= (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} e^{-ixt}f(x)\,dx \qquad (\mathrm{M}_2) \end{aligned}$$

The symbol (M_2) indicates that the integrals in equations (1) are to be taken in the sense (2) l. i. m. $T \to \infty \int_{-T}^{T}$. A general principle, due to N. Wiener, states that $f(x)$ and $\phi(t)$ cannot both be too small at infinity unless they are both zero almost everywhere. This principle was first realized as a theorem by G. H. Hardy. A short time later further theorems were obtained by G. W. Morgan. They proved somewhat more precise results which imply that if

$$|\phi(t)| = O(e^{-|t|^p}) \quad (t \to \pm\infty); \qquad |f(x)| = O(e^{-|x|^q}) \quad (x \to \pm\infty),$$

where p and q are positive and such that $p^{-1} + q^{-1} < 1$, then $f(x)$ and $\phi(t)$ are zero almost everywhere.

We shall consider the case where $\phi(t)$ approaches zero very slowly while $f(x)$ approaches zero very rapidly, a case which has applications to the theory of quasi-analytic functions. It is interesting to give a typical though very special case of our results, the general statement of which we postpone until Section 2. Let

$$(2) \qquad \phi(t)e^{|t|\theta(t)} \in L_2(-\infty, \infty),$$

where $\theta(t) = (\pi/2)[\log(|t| + e)]^{-1}$. If $f(x)$ is the L_2 Fourier transform of $\phi(t)$ and if

$$(3) \qquad f(x) = O(\exp\{-\exp\exp(1+\epsilon)|x|\}), \qquad (x \to +\infty \text{ or } x \to -\infty),$$

---

* Received March 22, 1949.

[1] Research supported in part by the Office of Naval Research.

for any $\epsilon > 0$, then $f(x) = 0$ almost everywhere. On the other hand, there exists a function $f(x) \not\equiv 0$, the L_2 transform of a function $\phi(t)$ satisfying (2), such that equation (3) holds for every $\epsilon < 0$. We may equally well treat one-sided conditions. Let $\phi(t)e^{|t|\theta(t)} \varepsilon L_2(-\infty, \infty)$, where

$$\theta(t) = \pi[\log(t+e)]^{-1} \qquad t > 0, \qquad \theta(t) = 0 \qquad t \leqq 0.$$

If $f(x)$ is the L_2 transform of $\phi(t)$ and if equation (3) holds for any $\epsilon > 0$, then $f(x)$ is zero almost everywhere.

Let $f(x)$ be an infinitely differentiable function defined for $-\infty < x < \infty$ such that

$$(4) \qquad |f^{(n)}(x)| \leqq A(2/\pi)^n n! \qquad (-\infty < x < \infty; n = 0, 1, \cdots).$$

$f(x)$ is then the restriction to the real axis of a function $f(z)$ analytic and bounded in every strip $|\operatorname{Im} z| \leqq l < \pi/2$. A simple application of the Phragmen-Lindelöf principle shows that if

$$(5) \qquad f(x) = O(\exp[-\exp(1+\epsilon)x]) \qquad (x \to +\infty)$$

for any $\epsilon > 0$, then $f(x) \equiv 0$. On the other hand there exists a function $f(x) \not\equiv 0$ satisfying (4) and such that equation (5) holds for every $\epsilon < 0$. Using our results on Fourier transforms we shall obtain similar results for quasi-analytic classes of functions defined by the more general inequalities

$$|f^{(n)}(x)| \leqq Ak^n M_n \qquad (-\infty < x < \infty; n = 0, 1, \cdots).$$

Let us consider a special case closely related to our previous examples. If

$$(6) \quad |f^{(n)}(x)| \leqq An!(\log(n+e))^n(2/\pi)^n \quad (-\infty < x < \infty; n = 0, 1, \cdots),$$

and if

$$(7) \qquad f(x) = O(\exp\{-\exp\exp(1+\epsilon)x\}) \qquad (x \to +\infty),$$

for $\epsilon > 0$, then $f(x) \equiv 0$. On the other hand, there exists a function $f(x) \not\equiv 0$, satisfying inequalities (6) and such that (7) holds for every $\epsilon < 0$.

2. The main theorem. We proceed immediately to the demonstration of our principal result.

THEOREM 2. *Let*

1. $\phi(t)e^{|t|\theta(t)} = \psi(t) \varepsilon L_2(-\infty, \infty)$
where $0 \leqq \theta(t) \leqq M$, $-\infty < t < \infty$, *for some* M;

2. $H(r) = \pi^{-1}\int_{-r}^{r} | t | \,\theta(t)\,[1+t^2]^{-1}dt$, $H(r) \to \infty$ *as* $r \to \infty$;

3. $f(x) = (2\pi)^{-1/2}\int_{-\infty}^{\infty} e^{ixt}\phi(t)\,dt$ $\qquad (M_2)$;

4. $f(x) = O(e^{-L(x)})$ $\qquad (x \to +\infty)$,

where for x *sufficiently large* $L(x)$ *is positive and* $L(x)/x$ *is strictly increasing to* $+\infty$. *If* $\overline{\lim\limits_{r\to\infty}}\, H[L(r)]r^{-1} > 1$, *then* $f(x) = 0$ *almost everywhere.*

We define

$$(1) \qquad F(w) = (2\pi)^{-1/2}\int_{-\infty}^{\infty} e^{-ixw}f(x)\,dx \qquad (w = u + iv).$$

If $v > 0$ then the integral (1) converges because of assumption 4. The function $F(w)$ is thus analytic for $v > 0$. Moreover since

$$\lim_{v\to 0+} \| e^{vx}f(x) - f(x) \|_2 = 0,$$

we have, by Parseval's theorem,

$$(2) \qquad \lim_{v\to 0+} \| F(u+iv) - \phi(u) \|_2 = 0.$$

Let S_r denote the strip $0 \leqq \operatorname{Im} w \leqq r$. The function $U(w)$ harmonic in S_r which assumes the boundary values $b_1(u)$ for $\operatorname{Im} w = 0$ and $b_2(u)$ for $\operatorname{Im} w = r$ is given by the formula

$$(3) \quad U(u+iv) = \int_{-\infty}^{\infty} b_1(u-t)k(v,t,r)\,dt + \int_{-\infty}^{\infty} b_2(u-t)k(r-v,t,r)\,dt,$$

where

$$(4) \quad k(v,t,r) = (2r)^{-1}\tan(\pi v/2r)[\tan^2(\pi v/2r) + \tanh^2(\pi t/2r)]^{-1}\cosh^{-2}(\pi t/2r).$$

This formula is easily verified by means of the mapping $\zeta = \tanh(\pi w/2r)$ which carries the strip S_r into the half plane $\operatorname{Im}\zeta \geqq 0$.

It is evident that the function $F(w)$ is bounded in every strip $\epsilon \leqq \operatorname{Im} w \leqq r$. Since $\log|F(w)|$ is a subharmonic function and since it is bounded from above we have, by the principle of majoration, that

$$\log|F(iv)| \leqq \int_{-\infty}^{\infty} \log|F(t+i\epsilon)|\, k(v-\epsilon, t, r-\epsilon)\,dt$$
$$\int_{-\infty}^{\infty} \log|F(t+ir)|\, k(r-v, t, r-\epsilon)\,dt.$$

Let us define $M(r) = \max\limits_{-\infty<u<\infty} \log|F(u+ir)|$. It can be verified that

$$\int_{-\infty}^{\infty} k(r - v, t, r - \epsilon)\,dt = (v - \epsilon)/(r - \epsilon).$$

Thus

$$\overline{\lim_{\epsilon\to 0+}} \int_{-\infty}^{\infty} \log |F(t + ir)|\, k(r - v, t, r - \epsilon)\,dt \leqq (v/r)M(r).$$

Equation (2) implies that $\lim_{\epsilon\to 0+} \| \log^{+} F(t + i\epsilon) - \log^{+} \phi(t) \|_2 = 0$, and thus that

$$\lim_{\epsilon\to 0+} \int_{-\infty}^{\infty} \log^{+} | F(t + i\epsilon) |\, k(v - \epsilon, t, r - \epsilon)\,dt = \int_{-\infty}^{\infty} \log^{+} | \phi(t) |\, k(v, t, r)\,dt.$$

Finally by what is essentially Fatou's theorem, see [10; p. 346],

$$\lim_{\epsilon\to 0+} \int_{-\infty}^{\infty} \log^{-} | F(t + i\epsilon) |\, k(v - \epsilon, t, r - \epsilon)\,dt \leqq \int_{-\infty}^{\infty} \log^{-} | \phi(t) |\, k(v, t, r)\,dt.$$

Combining these three equations we see that for $v > 0$

$$\begin{aligned} \log | F(iv) | &\leqq \int_{-\infty}^{\infty} \log | \phi(t) |\, k(v, t, r)\,dt + (v/r)M(r) \\ &\leqq -\int_{-\infty}^{\infty} | t \mid \theta(t) k(v, t, r)\,dt + \int_{-\infty}^{\infty} \log | \psi(t) |\, k(v, t, r)\,dt + (v/r)M(r). \end{aligned} \tag{5}$$

By $A(r) \sim B(r)$ we mean that $\lim_{r\to\infty} A(r)/B(r) = 1$.

LEMMA 2a.

$$\int_{-\infty}^{\infty} | t \mid \theta(t) k(v, t, r)\,dt \sim vH(r) \qquad (r \to +\infty).$$

We first assert that if $\epsilon > 0$ *then*

$$\int_{-\infty}^{\infty} | t \mid \theta(t) k(v, t, r)\,dt \sim \int_{-\epsilon r}^{\epsilon r} | t \mid \theta(t) k(v, t, r)\,dt \qquad (r \to +\infty). \tag{6}$$

This is an immediate consequence of the relation

$$\int_{|t| \geqq r\epsilon} | t \mid \theta(t) k(v, t, r)\,dt = O(1) \qquad (r \to +\infty), \tag{7}$$

which we will now prove. Recalling the definition of $k(v, t, r)$ and making the change of variable $t = rx$ we have

$$\int_{|t| \geqq r\epsilon} |t|\,\theta(t)k(v,t,r) = (r/2)\tan(\pi v/2r) \int_{|x| \geqq \epsilon} |x|\,\theta(rx)\,[\tan^2(\pi v/2r) + \tanh^2(\pi x/2)]^{-1}\cosh^{-2}(\pi x/2)\,dx \leqq \tfrac{1}{2}Mr\tan(\pi v/2r) \int_{|x| \geqq \epsilon} |x| \sinh^{-2}(\pi x/2)\,dx$$

where M is the constant of assumption 1. Since $\lim_{r \to +\infty} (r/2)\tan(\pi v/2r) = (\pi v/4)$, equation (7) follows.

If $\epsilon > 0$ then we may show by the same argument that

$$(8) \qquad (v/\pi)\int_{-r}^{r} |t|\,\theta(t)\,[v^2 + t^2]^{-1}dt \sim (v/\pi)\int_{-\epsilon r}^{\epsilon r} |t|\,\theta(t)\,[v^2 + t^2]^{-1}dt.$$

We next assert that given $\delta > 0$ there exists $\epsilon > 0$ and r_0 such that

$$(9) \qquad (1-\delta)(v/\pi)(v^2+t^2)^{-1} \leqq k(v,t,r) \leqq (1+\delta)(v/\pi)(v^2+t^2)^{-1} \qquad (|t| \leqq \epsilon r,\ r \geqq r_0).$$

This is easily seen. Combining equations (6), (8), and (9) we have proved that

$$\int_{-\infty}^{\infty} |t|\,\theta(t)k(v,t,r)\,dt \sim (v/\pi)\int_{-r}^{r} |t|\,\theta(t)\,[v^2+t^2]^{-1}dt.$$

It remains only to show that

$$(v/\pi)\int_{-r}^{r} |t|\,\theta(t)\,[v^2+t^2]^{-1}dt \sim vH(r).$$

This follows from the fact that

$$vH(r) - (v/\pi \int_{-r}^{r} |t|\,\theta(t)\,[v\ + t^2]^{-1}dt = (v/\pi)(v^2-1)\int_{-r}^{r} |t|\,\theta(t)\,[v^2+t^2]^{-1}[1+t^2]^{-1}dt = O(1) \qquad (r \to +\infty).$$

LEMMA 2b.

$$\overline{\lim_{r \to +\infty}} \int_{-\infty}^{\infty} k(v,t,r) \log |\psi(t)|\,dt < \infty.$$

We have

$$\int_{-\infty}^{\infty} k(v,t,r)\log|\psi(t)|\,dt \leqq \int_{-\infty}^{\infty} k(v,t,r)\log^{+}|\psi(t)|\,dt.$$

Since $\psi(t) \varepsilon L_2$, $\log^{+}|\psi(t)| \varepsilon L_2$, and hence

$$\overline{\lim_{r \to +\infty}} \int_{-\infty}^{\infty} k(v,t,r)\log|\psi(t)|\,dt \leqq (v/\pi)\int_{-\infty}^{\infty} \log^{+}|\psi(t)|\,[v^2+t^2]^{-1}dt.$$

For x and r sufficiently large, $x > x_0$, $r > r_0$, there is because of assumption 4 a unique solution $x = \Lambda(r)$ of the equation $xr = L(x)$.

LEMMA 2c.

$$M(r) \leqq \sim (r+1)\Lambda(r+1) \qquad (r \to +\infty).$$

We have

$$\log|F(u+ir)| \leqq \log[(\int_{-\infty}^{x_0} + \int_{x_0}^{\Lambda(r+1)} + \int_{\Lambda(r+1)}^{\infty}) e^{rx}|f(x)|\,dx]$$
$$\leqq \log[I_1(r) + I_2(r) + I_3(r)].$$

Now

$$I_1(r) = O(e^{x_0 r}) \qquad (r \to +\infty).$$

$$I_2(r) \leqq \int_{x_0}^{\Lambda(r+1)} e^{-x}e^{(r+1)x}dx \leqq e^{(r+1)\Lambda(r+1)} \qquad (r \geqq r_0).$$

$$I_3(r) \leqq \int_{\Lambda(r+1)}^{\infty} e^{-x}e^{(r+1)x}e^{-L(x)}dx = o(1) \qquad (r \to +\infty).$$

From these estimates our lemma follows immediately.

LEMMA 2d. *If* $0 < a$, $1 < b$, *then*

$$H(a) \geqq H(ab) - (2M/\pi)\log b, \tag{10}$$

$$H(r+a) \sim H(r) \qquad (r \to +\infty). \tag{11}$$

We have

$$H(ab) = \pi^{-1}\int_{-a}^{a}|t|\,\theta(t)[+t^2]^{-1}dt + \pi^{-1}\int_{a \leqq 1 \pm 1 \leqq ab}|t|\,\theta(t)[1+t^2]^{-1}dt$$
$$\leqq H(a) + (2M/\pi)\log b.$$

After transposition this is inequality (10). Setting $b = (r+a)/r$ we find that

$$H(r) \leqq H(r+a) \leqq H(r) + 2M/\pi)\log[(r+a)/r].$$

Equation (11) follows.

Using equation (5) and Lemmas 2a, 2b, and 2c, we see that, if for some $\epsilon > 0$

$$\lim_{r \to +\infty} r^{-1}[(r+1)\Lambda(r+1)] - (1-\epsilon)H(r) = -\infty,$$

or, what is the same thing, if

$$\overline{\lim_{r \to +\infty}}\, H(r)r/(r+1)\Lambda(r+1) > 1, \tag{12}$$

then $F(w) \equiv 0$ and hence $f(x) = 0$ almost everywhere. By equation (11) of Lemma 2d we see that

$$H(r)r/(r+1)\Lambda(r+1) \sim H(r+1)/\Lambda(r+1).$$

If we set $(r+1) = L(x)/x$, $\Lambda(r+1) = x$, we find that equation (12) is equivalent to

$$\overline{\lim_{x \to \infty}}\, H[L(x)/x]/x > 1.$$

By equation (10) of Lemma 2d we have $H[L(x)/x]/x \sim H[L(x)]/x$, and our theorem is proved.

3. The converse.

THEOREM 3. *If*

1. $0 \leqq \theta(t) \leqq M \quad (-\infty < t < \infty) \qquad \theta(t) = \theta(-t)$;
2. $t\theta(t) \varepsilon \uparrow \quad (0 \leqq t < \infty)$;
3. $H(r) = \pi^{-1} \int_{-r}^{r} |\, t \,|\, \theta(t)[1+t^2]^{-1} dt$, $H(r) \to \infty$ *as* $r \to +\infty$, *then*

there exists a function $\phi(t)$ *not equivalent to zero such that*

$$|\,\phi(t)\,|\, e^{|t|\theta(t)} \leqq [1+t^2]^{-1} \qquad (-\infty < t < \infty)$$

and such that

$$\text{if } f(x) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} e^{ixt}\phi(t)\,dt, \text{ then } f(x) = O(\exp - H^{-1}[(1-\epsilon)x]), \quad (x \to +\infty).$$

for every $\epsilon > 0$.

The equation

$$U(u+iv) = (v/\pi) \int_{-\infty}^{\infty} |\, t \,|\, \theta(t)\{[1+t^2]^{-1} - [v^2 + (t-u)^2]^{-1}\} dt$$

defines $U(u+iv)$ as a harmonic function in the half plane $v > 0$. For $v = 0$, U assumes the boundary values $|\, u \,|\, \theta(u)$, i.e., $\lim_{v \to 0+} U(u+iv) = -|\, u \,|\, \theta(u)$ almost everywhere.

We assert that

$$U(u+iv) \leqq U(iv) \qquad (-\infty < u < \infty; v > 0). \tag{1}$$

We have

$$\partial U(u+iv)/\partial u = v/\pi \int_{-\infty}^{\infty} |\, t \,|\, \theta(t)\{-2(t-u)[v^2+(t-u)^2]^{-2}\} dt$$
$$= v/\pi \int_{0}^{\infty} [|\, u+t \,|\, \theta(u+t) - |\, u-t \,|\, \theta(u-t)]\{-2t[v^2+t^2]^{-1}\} dt.$$

Using assumptions 1. and 2. we see that $\partial U/\partial u$ is negative for $u > 0$ and positive for $u < 0$, and this implies the validity of equation (1).

We will show that

$$U(iv)v^{-1} \sim H(v) \qquad (v \to +\infty). \tag{2}$$

Now $U(iv)v^{-1} = \pi^{-1}\int_{-\infty}^{\infty} |t|\,\theta(t)(v^2-1)(v^2+t^2)^{-1}(1+t^2)^{-1}dt$. If $\epsilon > 0$, then

$$U(iv)v^{-1} \sim \pi^{-1}\int_{-\epsilon v}^{\epsilon v} |t|\,\theta(t)(v^2-1)(v^2+t^2)^{-1}(1+t^2)^{-1}dt. \tag{3}$$

This is proved, as usual, by showing that

$$\pi^{-1}\int_{|t|\geqq \epsilon v} |t|\,\theta(t)(v^2-1)(v^2+t^2)^{-1}(1+t^2)^{-1}dt = O(1) \qquad (v \to +\infty).$$

Making the change of variable $t = vx$, we have

$$\begin{aligned}\pi^{-1}\int_{|t|\geqq \epsilon v} &|t|\,\theta(t)(v^2-1)(v^2+t^2)^{-1}dt \\ &= \pi^{-1}\int_{|x|\geqq \epsilon} |x|\,\theta(vx)(v^2-1)(1+x^2)^{-1}[1+v^2x^2]^{-1}dx \\ &\leqq M(\pi\epsilon)^{-1}\int_{|x|\geqq \epsilon}(1+x^2)^{-1}dx.\end{aligned}$$

Equation (3) now follows. Given $\delta > 0$ there exists $\epsilon > 0$ and $v_0 > 0$ such that

$$(1-\delta)(1+t^2)^{-1} \leqq (v^2-1)(v^2+t^2)^{-1}(1+t^2)^{-1} \leqq (1+\delta)(1+t^2)^{-2}$$
$$(|t| \leqq v\epsilon;\, v \geqq v_0).$$

The proof of equation (2) may now be completed exactly in the manner of Lemma 2a.

Let $V(u+iv)$ be a conjugate harmonic function of $U(u+iv)$. The function $\exp[U(w)+iV(w)]$ is analytic for $v > 0$ and is bounded in every strip $0 < v < r$. In particular its modulus does not exceed 1 in the strip $0 < v < 1$. We set

$$F(w) = (w+i)^{-2}\exp[U(w)+iV(w)],$$

and

$$f(x) = (2\pi)^{-\frac{1}{2}}e^{-\xi x}\int_{-\infty}^{\infty} e^{ixt}F(t+i\xi)\,dt \qquad (-\infty < x < \infty;\, \xi > 0).$$

It is easily verified using contour integration that $f(x)$ is independent of ξ.

By Fatou's theorem, see [1; vol. 2, p. 147), $\lim_{\epsilon\to 0+} F(t+i\xi)$ exists for almost all t. Call this limit $\phi(t)$. Since $|F(t+i\xi)| \leqq (1+t^2)^{-1}$ for $(0<\xi<1)$ we have, by the principle of dominated convergence,

$$f(x) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} e^{ixt}\phi(t)\,dt.$$

Moreover $|\phi(t)| \leqq e^{-|t|\theta(t)}(1+t^2)^{-1}$ almost everywhere.

It remains to prove that

$$f(x) = O(\exp\{-H^{-1}[(1-\epsilon)x]\}) \qquad (x\to+\infty)$$

for every $\epsilon>0$. From the definition of $f(x)$ and from equation (1) we see that

$$|f(x)| \leqq 2\pi)^{-1/2} e^{-\xi x} e^{U(i\xi)} \int_{-\infty}^{\infty} [(\xi+1)^2+t^2]^{-1} dt.$$

By equation (2), if ξ is sufficiently large, $|f(x)| \leqq \exp\{-\xi x + (1+\epsilon)\xi H(\xi)\}$. We now set $\xi = H^{-1}[(x-1)/(1+\epsilon)]$ to obtain, for x sufficiently large, $|f(x)| \leqq \exp\{-H^{-1}[(x-1)/(1+\epsilon)]\}$. Since, when x is large, $x(1-\epsilon) \leqq (x-1)/(1+\epsilon)$, we have $|f(x)| \leqq \exp\{-H^{-1}[(1-\epsilon)x]$, and this completes the proof of Theorem 3.

4. Quasi-analytic functions. Let $C\{M_n\}$ denote the class of functions $f(x)$ defined and infinitely differentiable for $-\infty<x<\infty$ and such that

$$(1) \qquad |f^{(n)}(x)| \leqq Ak^nM_n \qquad (-\infty<x<\infty;\, n=0,1,\cdots),$$

where A and k are constants which may depend upon $f(x)$. The convex regularization $\{M_n{}^c\}$ of the sequence $\{M_n\}$ is defined by the equations

$$(2) \qquad T(r) = \operatorname*{Max}_{n\geqq 0}\,(r^n/M_n), \qquad M_n{}^c = \operatorname*{Max}_{r\geqq 0}\,(r^n/T(r)).$$

See [8]. It may be shown that

$$(3) \qquad M_n{}^c \leqq M_n, \qquad M_n{}^c \leqq (M_{n-j}{}^c)^{k/(j+k)}(M_{n+k}{}^c)^{j/(j+k)}.$$

The class $C\{M_n\}$ is said to be quasi-analytic if a function $f(x) \in C\{M_n\}$ which vanishes with all its derivatives at a point $x=x_0$ is necessarily indentically zero. It is well known, see [8], that a necessary and sufficient condition for a class $C\{M_n\}$ to be quasi-analytic is that the integral

$$\int^{\infty} \log T(r)\, r^{-2} dr$$

diverge.

The class $C\{M_n\}$ is said to contain the class $C\{n!\}$, the analytic class, if every $f(x)$ in $C\{n!\}$ belongs also to $C\{M_n\}$. Using the function $f(x) = (i + x)^{-1}$ we see that a necessary and sufficient condition for $C\{M_n\}$ to contain $C\{n!\}$ is that

$$n! \leqq Bb^n M_n \qquad (n = 0, 1, \cdots), \tag{4}$$

for some positive constants B and b.

Kolmogoroff, [5], has shown that, if $m_n = \operatorname*{Sup}_{-\infty < x < \infty} |f^{(n)}(x)|$,

$$m_n \leqq (\pi/2)(m_{n-j})^{k/(j+k)}(m_{n+k})^{j/(j+k)}.$$

It follows that if $f(x)$ satisfies inequalities (1) then

$$|f^{(n)}(x)| \leqq (\tfrac{1}{2}\pi A) k^n M_n{}^c \qquad (-\infty < x < \infty; n = 0, 1, \cdots). \tag{5}$$

THEOREM 4. *Let*

1. $C\{M_n\} \supset C\{n!\}$;
2. $H(r) = (2/\pi)\int^r \log T(u) u^{-2} du$, $H(r) \to \infty$ *as* $r \to +\infty$;
3. $|f^{(n)}(x)| \leqq Ak^n M_n \qquad (-\infty < x < \infty; n = 0, 1, \cdots)$.

If $f(x) = O(e^{-L(x)})$, $(x \to +\infty)$, *where* $L(x)$ *is positive and strictly increasing to infinity, and if* $\overline{\lim}_{x \to +\infty} H[L(x)]/x > k$, *then* $f(x) \equiv 0$.

Conversely, if assumptions 1. *and* 2. *are satisfied, then there exists a function* $f(x)$ *satisfying* 3. *and such that* $f(x) = O(\exp - H^{-1}(k'x))$, $(x \to +\infty)$ *for every* $k' < k$.

We define $F(x) = (\sin x/x) f(x)$. We assert that if $k_1 > k$ then there exists a constant A_1 such that

$$\| F^{(n)}(x) \|_2 \leqq A_1 k_1{}^n M_n{}^c \qquad (n = 0, 1, \cdots). \tag{6}$$

We have $F^{(n)}(x) = \sum_{j=0}^{n} \binom{n}{j} f^{(n-j)}(x)(\sin x/x)^{(j)}$. Using equation (5) we obtain

$$\| F^{(n)}(x) \|_2 \leqq \sum_{j=0}^{n} \binom{n}{j} \tfrac{1}{2}\pi A k^{n-j} M_{n-j}{}^c \, \| (\sin x/x)^{(j)} \|_2.$$

Now

$$(\sin x/x)^{(j)} = \tfrac{1}{2}\int_{-1}^{1} e^{ixt}(it)^j dt, \qquad \| (\sin x/x)^{(j)} \|_2 = (\pi/2)^{1/2} [\int_{-1}^{1} t^{2j} dt]^{1/2}$$
$$\leqq (\pi/2)^{1/2} \qquad (j = 0, 1, \cdots).$$

Also, by equation (3), $M_{n-j}^c \leqq (M_n^c)^{1-(j/n)}(M_0^c)^{j/n}$. Thus

$$\begin{aligned}\| F^{(n)}(x) \|_2 &\leqq (\pi/2)^{3/2} A \sum_{j=0}^{n} \binom{n}{j} k^{n-j} (M_n^c)^{1-(j/n)} (M_0^c)^{j/n} \\ &\leqq (\pi/^{3/2} A [(M_n^c)^{1/n} k + (M_0^c)^{1/n}]^n \\ &\leqq (\pi/2)^{3/2} A k^n M_n^c [1 + (M_0^c/M_n^c)^{1/n} k^{-1}]^n.\end{aligned}$$

Since, from assumption 1., $(M_n^c)^{1/n} \to \infty$ as $n \to \infty$, equation (6) follows.

Because $F(x) \in L_2(-\infty, \infty)$, there is a function $\phi(t) \in L_2(-\infty, \infty)$ such that

$$F(x) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} e^{ixt} \phi(t)\, dt \qquad (\mathrm{M}_2).$$

Moreover,

$$F^{(n)}(x) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} e^{ixt} (it)^n \phi(t)\, dt \qquad (\mathrm{M}_2).$$

This implies that if $k_2 > k_1$, then

$$\| \sum_{n=0}^{\infty} (t/k_2)^n M_n^{-1} \phi(t) \|_2 \leqq \sum_{n=0}^{\infty} k_2^{-n} M_n^{-1} \| t^n \phi(t) \|_2 = \sum_{n=0}^{\infty} k_2^{-n} M_n^{-1} \| F^{(n)}(x) \|_2$$

$\leqq A_1 \sum_{n=0}^{\infty} (k_1/k_2)^n < \infty$. This in turn implies that $\| T(|t|/k_2)\phi(t) \|_2 < \infty$.

Let us define $\theta(t)$ by the equations

$$\theta(t) = 0, \quad (|t| \leqq 1); \qquad |t|\,\theta(t) = \log^+[T(|t|/k_2)], \quad (|t| > 1).$$

We then have $e^{|t|\theta(t)}\phi(t) \in L_2(-\infty, \infty)$. Clearly $\theta(t) \geqq 0$. Assumption 1. implies inequality (4) for suitably chosen B and b, and equation (4) implies that $\theta(t) \leqq M$ for a suitably chosen constant M. Finally $F(x) = O(e^{-L(x)})$, $(x \to +\infty)$. Applying Theorem 2 we see that if

$$\overline{\lim_{x \to +\infty}}\, H^*[L(x)]/x > 1, \text{ where } H^*(x) = (2/\pi) \int^{r} \log^+[T(t/k\ )] t^{-2} dt,$$

then $f(x) \equiv 0$. Making the change of variable $t = k_2 u$ in equation (6) we obtain

$$H^*(r) = (2/\pi k_2) \int^{k_2 r} \log^+ T(u) u^{-2} du \sim k_2^{-1} H(r) \qquad (r \to +\infty),$$

by Lemma 2d. Since k_2 may be taken as near k as we please the first part of our theorem follows.

By Theorem 3 there exists a function $\phi(t)$ such that

$$|\phi(t)| \leqq e^{-\log^+ T(|t|)}(1+t^2)^{-1} \qquad (-\infty < t < \infty),$$

and such that if

$$f(x) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} e^{ixt}\phi(t)\,dt, \text{ then } f(x) = O\{-H^{-1}[(1-\epsilon)x]\}) \qquad (x \to +\infty)$$

for every $\epsilon > 0$. Now $f^{(n)}(x) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} e^{ixt}(it)^n\phi(t)\,dt$,

$$|f^{(n)}(x)| \leqq (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} t^n[T(t)]^{-1}(1+t^2)^{-1}dt$$
$$\leqq (\pi/2)^{\frac{1}{2}} \operatorname{Max} t^n/T(t) \leqq (\pi/2)^{\frac{1}{2}} M_n^c \leqq (\pi/2)^{\frac{1}{2}} M_n.$$

This completes the proof for $k=1$. For general k we need only consider $f(kx)$.

Let us agree to write $l_1(x) = \log x$, $l_2(x) = \log\log x, \cdots$, and also $e_1(x) = \exp x$, $e_2(x) = \exp\exp x, \cdots$. The following lemma is well known, but since I have been unable to find a reference the proof is included.

LEMMA 4. *If*

$$1.\quad N_n^{(m)} = \begin{cases} n![l_1(n)l_2(n)\cdots l_m(n)]^n & n > e_m(1) \\ n! & n \leqq e_m(1) \end{cases}$$

$$2.\quad T_m(x) = \operatorname*{Max}_{n \geqq 0} x^m/N_n^{(m)},$$

then $\log T_m(x) \sim x[l_1(x)l_2(x)\cdots l_m(x)]^{-1}$, $(x \to +\infty)$.

Let

$$(10)\quad \log T^*_m(x) = \operatorname*{Max}_{n \geqq 0}[n\log x + n - nl_1(n) - nl_2(n) - \cdots - nl_m(n)].$$

Using Stirling's formula it is easily seen that if $0 < \epsilon < 1$, then, for x sufficiently large,

$$\log T^*_m((1-\epsilon)x) \leqq \log T_m(x) \leqq \log T^*_m||1+\epsilon)x),$$

so that it is sufficient to prove

$$\log T^*_m(x) \sim x[l_1(x)l_2(x)\cdots l_m(x)]^{-1}.$$

Let $n[x]$ be the integer for which the maximum in equation (10) is attained;

then $n[x]-1<n(x)<n[x]+1$, where $n(x)$ is the solution of the equation

$$(d/dn)\,[n\log x+n-nl_1(n)-\cdots-nl_m(n)]=0,$$

which we may rewrite as $\log x-L(n)=0$, where

$$L(n)=l_1(n)+l_2(n)-\cdots+l_m(n)+[l_1(n)]^{-1}+[l_1(n)l_2(n)]^{-1}$$
$$+\cdots+[l_1(n)l_2(n)\cdots l_m(n)]^{-1}.$$

Clearly

$$\log x=l_1(n[x])+\cdots+l_m(n[x])+o(1) \qquad (x\to+\infty).$$

Inserting this in equation (10), we find that

$$\log T^*_m(x)=n[x][1+o(1)]\sim n(x) \qquad (x\to+\infty).$$

It may be verified that if $0<\epsilon<1$, and if x is sufficiently large, then $L(n)>\log x$ when $n=(1+\epsilon)x[l_1(x)l_2(x)\cdots l_m(x)]^{-1}$, and $L(n)<\log x$ when $n=(1-\epsilon)x[l_1(x)l_2(x)\cdots l_m(x)]^{-1}$. It follows that

$$n(x)\sim x[l_1(x)l_2(x)\cdots l_m(x)]^{-1} \qquad (x\to+\infty), \tag{11}$$

and our lemma is proved.

Let us set $H_m(r)=(2/\pi)\int^r \log T_m(u)u^{-2}du.$

By Lemma 4 we have

$$H_m(r)\sim(2/\pi)l_{m+1}(r) \qquad (r\to+\infty). \tag{12}$$

Combining equation (12) with Theorem 4 we obtain

COROLLARY 4. *Let*

$$N_n^{(m)}=\begin{cases} n![l_1(n)l_2(n)\cdots l_m(n)]^n & n>e_m(1) \\ n! & n\leqq e_m(1) \end{cases}.$$

If

$$|f^{(n)}(x)|\leqq AN_n^{(m)}(2/\pi)^n \qquad (-\infty<x<\infty;\, n=0,1,\cdots), \tag{13}$$

and if

$$f(x)=O(e_1\{-e_{m+1}[(1+\epsilon)x]\}) \qquad (x\to+\infty) \tag{14}$$

for $\epsilon>0$, then $f(x)\equiv0$. On the other hand there exists a function

$f(x) \not\equiv 0$, satisfying inequalities (13) and such that relation (14) holds for every $\epsilon < 0$.

HARVARD UNIVERSITY AND WASHINGTON UNIVERSITY.

BIBLIOGRAPHY.

1. L. Bieberbach, *Lehrbuch der Funktionentheorie*, Leipzig, 1931.

2. A. Gorny, "Contribution à l'étude des fonctions dérivables d'une variable réelle," *Acta Mathematica*, vol. 71 (1939), pp. 317-358.

3. G. H. Hardy, "A theorem concerning Fourier transforms," *Journal of the London Mathematical Society*, vol. 8 (1933), pp. 227-231.

4. I. I. Hirschman, Jr., "The behaviour at infinity of certain convolution transforms," submitted to the *Transactions of the American Mathematical Society*.

5. A. Kolmogoroff, "Une généralization de l'inégalité de M. J. Hadamard entre les bornes supérieures des dérivées successives d'une fonction," *Comptes Rendus*, vol. 207 (1938), p. 764.

6. A. E. Ingham, "A note on Fourier transforms," *Journal of the London Mathematical Society*, vol. 9 (1934), pp. 29-32.

7. N. Levinson, *Gap and density theorems*, New York, 1940, pp. 73-87.

8. S. Mandelbrojt, "Analytic functions and classes of infinitely differentiable functions," *The Rice Institute Pamphlet*, vol. 29 (1942), pp. 1-142.

9. G. W. Morgan, "A note on Fourier transforms," *Journal of the London Mathematical Society*, vol. 9 (1934), pp. 187-192.

10. E. C. Titchmarsh, *The theory of functions*, Oxford, 1932.

THE MARRIAGE PROBLEM.*

By Paul R. Halmos and Herbert E. Vaughan.

In a recent issue of this journal Weyl[1] proved a combinatorial lemma which was apparently considered first by P. Hall.[2] Subsequently Everett and Whaples[3] published another proof and a generalization of the same lemma. Their proof of the generalization appears to duplicate the usual proof of Tychonoff's theorem.[4] The purpose of this note is to simplify the presentation by employing the statement rather than the proof of that result. At the same time we present a somewhat simpler proof of the original Hall lemma.

Suppose that each of a (possibly infinite) set of boys is acquainted with a finite set of girls. Under what conditions is it possible for each boy to marry one of his acquaintances? It is clearly necessary that every finite set of k boys be, collectively, acquainted with at least k girls; the Everett-Whaples result is that this condition is also sufficient.

We treat first the case (considered by Hall) in which the number of boys is finite, say n, and proceed by induction. For $n = 1$ the result is trivial. If $n > 1$ and if it happens that every set of k boys, $1 \leqq k < n$, has at least $k + 1$ acquaintances, then an arbitrary one of the boys may marry any one of his acquaintances and refer the others to the induction hypothesis. If, on the other hand, some group of k boys, $1 \leqq k < n$, has exactly k acquaintances, then this set of k may be married off by induction and, we assert, the remaining $n - k$ boys satisfy the necessary condition with respect to the as yet unmarried girls. Indeed if $1 \leqq h \leqq n - k$, and if some set of h bachelors were to know fewer than h spinsters, then this set of h bachelors together with the k married men would have known fewer than $k + h$ girls. An

---

* Received June 6, 1949.

[1] H. Weyl, "Almost periodic invariant vector sets in a metric vector space," *American Journal of Mathematics*, vol. 71 (1949), pp. 178-205.

[2] P. Hall, "On representation of subsets," *Journal of the London Mathematical Society*, vol. 10 (1935), pp. 26-30.

[3] C. J. Everett and G. Whaples, "Representations of sequences of sets," *American Journal of Mathematics*, vol. 71 (1949), pp. 287-293. Cf. also M. Hall, "Distinct representatives of subsets," *Bulletin of the American Mathematical Society*, vol. 54 (1948), pp. 922-926.

[4] C. Chevalley and O. Frink, Jr., "Bicompactness of Cartesian products," *Bulletin of the American Mathematical Society*, vol. 47 (1941), pp. 612-614.

application of the induction hypothesis to the $n-k$ bachelors concludes the proof in the finite case.

If the set B of boys is infinite, consider for each b in B the set $G(b)$ of his acquaintances, topologized by the discrete topology, so that $G(b)$ is a compact Hausdorff space. Write G for the topological Cartesian product of all $G(b)$; by Tychonoff's theorem G is compact. If $\{b_1,\cdots,b_n\}$ is any finite set of boys, consider the set H of all those elements $g=g(b)$ of G for which $g(b_i)\neq g(b_j)$ whenever $b_i\neq b_j$, $i,j=1,\cdots,n$. The set H is a closed subset of G and, by the result for the finite case, H is not empty. Since a finite union of finite sets is finite, it follows that the class of all sets such as H has the finite intersection property and, consequently, has a non empty intersection. Since an element $g=g(b)$ in this intersection is such that $g(b')\neq g(b'')$ whenever $b'\neq b''$, the proof is complete.

It is perhaps worth remarking that this theorem furnishes the solution of the celebrated problem of the monks.[5] Without entering into the history of this well-known problem, we state it and its solution in the language of the preceding discussion. A necessary and sufficient condition that each boy b may establish a harem consisting of $n(b)$ of his acqaintances, $n(b)=1, 2, 3,\cdots$, is that, for every finite subset B_0 of B, the total number of acquaintances of the members of B_0 be at least equal to $\Sigma n(b)$, where the summation runs over every b in B_0. The proof of this seemingly more general assertion may be based on the device of replacing each b in B by $n(b)$ replicas seeking conventional marriages, with the understanding that each replica of b is acquainted with exactly the same girls as b. Since the stated restriction on the function n implies that the replicas satisfy the Hall condition, an application of the Everett-Whaples theorem yields the desired result.

UNIVERSITY OF CHICAGO
AND
UNIVERSITY OF ILLINOIS.

---

[5] H. Balzac, *Les Cent Contes Drôlatiques*, IV, 9: *Des moines et novices*, Paris (1849).

NOTE ON A RESULT OF L. FUCHS ON ORDERED GROUPS.*[1]

By C. J. Everett.

Let G be an abelian group. If it admits a linear order, every non-zero element satisfies the relation $a > 0$ or $a < 0$, whence $na > 0$ or $na < 0$ for all $n = 1, 2, \cdots$. Hence, G has the property $(*)$: every non-zero element of G is of infinite order.

If P_0 is an arbitrary partial order on an abelian group G of type $(*)$, it possesses a linear extension L. To see this, it is convenient to recall that a partial order on G is completely defined by its set N of elements $p \geqq 0$. The latter has the characterizing properties: A) N is closed under addition, B) contains zero, C) contains no element along with its inverse except zero. Note that if positive multiples nx and $m(-x)$ are in N, then[2] mnx and $nm(-x) = -(mnx)$ are in N and $x = 0$.

If N is the non-negative set of a partial order P and neither x nor $-x$ is in N, three mutually exclusive cases may obtain: 1) some positive multiple nx is in N, 2) some positive multiple $m(-x)$ is in N, 3) no positive multiple nx nor $m(-x)$ is in N. In case 1, define N^* as the set of all elements of form $p + nx$, p in N, $n = 0, 1, 2, \cdots$; in case 2, similarly but with x replaced by $-x$; in case 3, in either way. It is trivial that N^* satisfies A, B, C, and contains N properly.

The class of all sets N satisfying A, B, C, and containing the set N_0 of the original P_0, is a partially ordered set under set inclusion; and every linearly ordered subclass has an upper bound, namely its set union. Hence there are maximal sets containing N_0. But a maximal set must contain either x or $-x$ for every x of G, and the corresponding order is a linear extension of P_0.

Los Alamos Scientific Laboratory.

---

* Received March 16, 1949.

[1] L. Fuchs, "On the extension of the partial order of groups," *American Journal of Mathematics*, vol. 72 (1950), pp. 191-194.

[2] Remark by L. Fuchs simplifying original argument.

THE GAUSS TEST FOR RELATIVE CONVERGENCE.*

By R. L. Goodstein.

Our problem is to establish in extended *recursive number theory* a test for *relative convergence* analogous to the familiar Gauss test for positive series. Recursive number theory is a formal system admitting only (primitive or multiply) recursive functions or relations (and therefore only *bounded* existential and universal operators). Formulations of recursive number theory have been given, *inter alia*, by [1], [2] and [3][1]; we consider a simple extension of recursive number theory to rational functions of positive integral arguments. Propositions and proofs of the extended system R may be translated into the parent system with the help of the familiar definitions of fractions in terms of pairs of natural numbers and of rational numbers in terms of pairs of fractions. A *fractional recursive* function $f(n_1, n_2, \cdots, n_k)$ is a pair of recursive functions $a(n_1, n_2, \cdots, n_k)/b(n_1, n_2, \cdots, n_k)$ such that $b(n_1, n_2, \cdots, n_k) > 0$ for all natural numbers $n_1, n_2, \cdots, n_k$ and $a/b \gtreqless a'/b'$ according as $ab' \gtreqless a'b$, and $a/b + a'/b' = (ab' + a'b)/bb'$, $(a/b) \cdot (a'/b') = aa'/bb'$, and a *rational recursive* function $r(n_1, n_2, \cdots, n_k)$ is a pair of fractional recursive functions $[f(n_1, n_2, \cdots, n_k), g(n_1, n_2, \cdots, n_k)]$ such that $[f, g] \gtreqless [f', g']$ according as $f + g' \gtreqless f' + g$ and $[f, g] + [f', g'] = [f + f', g + g']$, $[f, g] \cdot [f', g'] = [ff' + gg', fg' + f'g]$, etc.

The concept of *relative convergence* was introduced in [2] to serve for the expression in a rational field of an analogue of a convergent sequence of real numbers. A rational recursive function $f(m, n)$ is said to be convergent in m relative to n if (roughly speaking) $f(m, n)$ satisfies the Cauchy criterion for convergence in m only for sufficiently great values of n; convergence relative to n does not involve convergence for any fixed value of n, since for instance m/n is convergent in m relative to n, but m/n is not convergent for any given value of n.

Definitions in R. A rational recursive function $s(n)$ is said to be convergent if it is associated with a recursive function $n(k)$ such that $\{n \geqq n(k)\} \rightarrow | s(n) - s(n(k)) | < 1/(k + 1)$, where n, k are free variables. A rational number p/q is said to be the limit of $s(n)$ (and $s(n)$ is said to tend to p/q,

---

* Received July 30, 1948; revised December 27, 1949.

[1] The numbers in square brackets refer to the bibliography at the end of the paper.

denoted by $s(n) \rightarrow p/q$) if $s(n)$ is associated with a recursive function $n(k)$ such that $\{n \geqq n(k)\} \rightarrow |\,p/q - s(n)\,| < 1/(k+1)$, where n, k are free variables. *If* $s(n) \rightarrow p/q$ then $s(n)$ is convergent, for $|\,s(n) - s(n(k))\,| \leqq |\,s(n) - p/q\,| + |\,s(n(k)) - p/q\,| < 2/(k+1)$, when $n \geqq n(k)$. The function $s(n)$ is *divergent* if it is associated with a recursive $d(n)$ and a constant $k \geqq 1$ such that $|\,s(n + d(n)) - s(n)\,| \geqq 1/k$, where n is a free variable.

The function $s(n)$ is said *not* to tend to a limit if there are recursive functions $\nu(p, q) \geqq 1$, $n(p, q, r) \geqq r$, such that

$$\{n = n(p, q, r)\} \rightarrow |\,s(n) - p/(q+1)\,| \geqq 1/\nu(p, q),$$

where p, q and r are free variables.

In a rational field, of course, a convergent function does not necessarily tend to a limit. For instance, if $s(n) = u(n)/n!$, where $u(0) = 1$, $u(n+1) = (n+1)u(n) + 1$ then $s(n)$ converges, for if $n > q$ then $0 < s(n) - s(q) < 1/q(q!)$; but

$$\{s(q) - p/q\}(q!) = u(q) - p\{(q-1)!\}$$
$$= (q-1)[q \cdot u(q-2) + 1 - p\{(q-2)!\}] + 2$$

$= 2 (\mathrm{mod}\ (q-1))$, $q \geqq 4$. Hence $|\,s(n) - p/q\,| \geqq (1 - 1/q)/q!$, for $n > q \geqq 4$; in the cases $q = 1, 2, 3$, for $n > 4$ the difference $|\,s(n) - p/q\,|$ is least for $p = 8$, $q = 3$ and $|\,s(n) - 8/3\,| > 1/24$. Thus $s(n)$ does not tend to p/q for any p, q.

The restriction to bounded operators does not seriously complicate the theory of ordinary convergence. The principal classical result lost by this restriction is that which asserts the convergence of a bounded monotonic sequence. For a proof in R that any non-decreasing recursive function $f(n)$ is associated with a recursive $n_f(k)$ such that

$$n \geqq n_f(k) \rightarrow |\,f(n) - f(n_f(k))\,| < 1/(k+1)$$

entails that an equivalent of the formula

$$(n)(k)\{n \geqq n_f(k) \rightarrow |\,f(n) - f(n_f(k))\,| < 1/(k+1)\}$$

is provable in the system Z_μ of Hilbert-Bernays [3], and so if $T(m)$ is a recursive relation such that $T(v)$ holds for any assigned numeral v but $(m)T(m)$ is non-demonstrable in Z_μ, and if $f(n)$ is the characteristic function of the recursive relation $(m)\{m \leqq n \,\&\, T(m)\}$, then $f(n)$ is demonstrably recursive, non-decreasing and bounded by unity, in Z_μ. By hypothesis there is a recursive function $n_f(x)$ in R and so a recursive term $n_f(x)$ in Z_μ such that the formula

$$(n)\{n \geqq n_f(0) \rightarrow |f(n) - f(n_f(0))| < 1\}$$

is demonstrable in Z_μ, whence it follows that

$$(n)\{n \geqq n_f(0) \rightarrow f(n) = f(n_f(0))\}$$

is demonstrable. Since $n_f(x)$ is primitive recursive, there is a numeral w such that the equation $n_f(0) = w$ is provable in Z_μ; but $T(r)$ is provable in Z_μ for $r = 1, 2, \cdots, w$, so that $(n)(n \leqq w \,\&\, f(n) = 0)$ is provable, and in particular $f(n_f(0)) = 0$ is provable. It follows that $(n)(f(n) = 0)$ is demonstrable in Z_μ, from which we conclude that $(m)T(m)$ is demonstrable in Z_μ, in contradiction to the hypothesis on $T(m)$.

We preface our discussion of relative convergence by a brief examination of the validity in R of the elementary tests for ordinary convergence of positive series. We denote by $a(n)$, $u(n)$, etc., positive (non-zero) recursive functions, and define for any function $f(n)$,

$$\sum_{r=m}^{m+0} f(r) = f(m), \quad \sum_{r=m}^{m+n+1} f(r) = \sum_{r=m}^{m+n} f(r) + f(m+n+1),$$

so that, e. g., $\sum_{r=m}^{m+n} a(r)$ is monotonic increasing with n.

If x is a rational variable then $\sum_{r=0}^{n} x^r$ is convergent for $0 < x < 1$, and divergent for $x \geqq 1$; for if $x \neq 1$, $\sum_{r=0}^{n-1} x^r = (1 - x^n)/(1 - x)$ (proof by induction), and if $y > 0$, $(1+y)^n \geqq 1 + ny$ (proof by induction), so that $1/(1+y)^n \rightarrow 0$, whence, taking $y = (1/x) - 1 > 0$, $\sum_{r=0}^{n-1} x^r \rightarrow 1/(1-x)$, provided that $0 < x < 1$. If $x \geqq 1$ then $\sum_{r=0}^{n+1} x^r - \sum_{r=0}^{n} x^r = x^{r+1} \geqq 1$, so that $\sum_{r=0}^{n} x^r$ diverges. Similarly if $a(n)$ does not tend to zero then $\sum_{r=0}^{n} a(r)$ diverges, for $\sum_{r=0}^{n+1} a(r) - \sum_{r=0}^{n} a(r) = a(n+1) \geqq 1/\nu(0,0)$ provided that $n + 1 = n(0, 0, p)$.

If $\sum_{r=0}^{n} a(r)$ diverges then $\{1/\sum_{r=m}^{n} a(r)\} \rightarrow 0$ for any fixed m; let $s(n) = \sum_{r=m}^{m+n} a(r)$ and $\phi(0) = 0$, $\phi(n+1) = \phi(n) + d(\phi(n))$, so that $s(\phi(n+1)) - s(\phi(n)) \geqq 1/k$, by hypothesis, whence $s(\phi(n)) \geqq n/k$ and so, for $p \geqq \phi(n)$, $1/s(p) \leqq 1/s(\phi(n)) \leqq k/n \rightarrow 0$.

The elementary tests.

The comparison tests. (i) If $u(n)$, $v(n)$ are associated with a non-zero constant k, such that $u(n) \leqq kv(n)$, then the convergence of $\sum_{r=0}^{n} v(r)$ entails that of $\sum_{r=0}^{n} u(r)$, and $\sum_{r=0}^{n} v(r)$ diverges with $\sum_{r=0}^{n} u(r)$. For $\sum_{r=m}^{m+n} u(r) \leqq k \sum_{r=m}^{m+n} v(r)$.

(ii) If $u(n+1)/u(n) \leqq v(n+1)/v(n)$, then $\sum_{r=0}^{n} u(r)$ converges with $\sum_{r=0}^{n} v(r)$ and the latter diverges with the former. For $u(n) \leqq \{u(0)/v(0)\}v(n)$.

Cauchy's first test. If for some constant k, $u(n) \leqq k^n < 1$, then $\sum_{r=0}^{n} u(r)$ converges (by the foregoing test); if $u(n) \geqq 1$, for $n = n(r)$, where $n(r+1) > n(r)$, then $\sum_{r=0}^{n} u(r)$ diverges (for $u(n)$ does not tend to zero).

D'Alembert's test. If for some k, $u(n+1)/u(n) \leqq k < 1$, then $u(n+1)/u(n) \leqq k^{n+1}/k^n$, and so $\sum_{r=0}^{n} u(r)$ converges, by (ii) above. If $u(n+1)/n(n) \geqq 1$, then $\sum_{r=0}^{n} u(r)$ diverges, since $\sum_{r=0}^{n} 1$ diverges.

Kummer's test (original form). If for some constant β

$$c(n)\{u(n)/u(n+1)\} - c(n+1) \geqq \beta > 0,$$

and $c(n)u(n) \to 0$, then $\sum_{r=0}^{n} u(r)$ converges, since

$$\beta \sum_{r=n+1}^{N} u(r) \leqq c(n)u(n) - c(N)u(N) < c(n)u(n) \to 0.$$

Dini's form of Kummer's test, which omits the condition $c(n)u(n) \to 0$, is not valid in R (unless some additional condition is imposed, for instance that $\sum 1/c(n)$ converges, when convergence follows by the comparison test). In fact we can show that Dini's form of the test is equivalent to the assertion of the convergence of any bounded monotonic sequence. For if $s(n)$ is any strictly increasing sequence bounded by B and if $u(n) = s(n) - s(n-1)$, $v(n) = B - \sum_{r=1}^{n} u(r)$ and $c(n) = v(n)/u(n)$, then $c(n)\{(n)/u(n+1)\} - c(n+1) = 1$.

Cauchy's condensation test. If $f(n)$ is steadily decreasing, then $\sum_{r=0}^{n} f(r)$ and $\sum_{r=0}^{n} 2^r f(2^r)$ converge and diverge together. For $\sum_{r=2^n}^{2^{n+1}-1} f(r) < 2^n f(2^n)$ and therefore $\sum_{r=2^n}^{2^{N+1}-1} f(r) < \sum_{r=n}^{N} 2^r f(2^r)$, whence $\sum_{r=2^n}^{M} f(r) < \sum_{r=n}^{N} 2^r f(2^r)$, if $2^{N+1} - 1 \geqq M$.

Furthermore $\sum_{r=2^n+1}^{2^{n+1}} f(r) \geqq 2^n f(2^{n+1})$ and so $\sum_{r=m}^{2^{N+1}} f(r) \geqq \sum_{r=n}^{N} 2^r f(2^{r+1})$, if $2^n + 1 \geqq m$. It follows that $\sum_{r=1}^{n} 1/r$ is divergent, and therefore $\{1/\sum_{r=m}^{m+n} (1/r)\} \to 0$, for fixed m. Also, $\sum_{r=1}^{n} 1/r^2$ converges since $\sum_{r=1}^{n} 1/2^r$ converges.

Raabe's tests. (i) If $u(n)/u(n+1) \geqq 1 + k/n$, for some $k > 1$, then $\sum_{r=0}^{n} u(r)$ converges.

We observe first that $n \cdot u(n)/(n+1) \cdot u(n+1) \geqq 1 + (k-1)/(n+1)$ and so, since

$$(1 + (k-1)/(n+1))(1 + (k-1)/(n+2)) \cdots$$
$$\cdots (1 + (k-1)/N) > (k-1) \sum_{r=n+1}^{N} 1/r,$$

therefore $N \cdot u(N) \leqq n \cdot u(n)/(k-1) \sum_{r=n+1}^{N} (1/r) \to 0$, keeping n fixed. Hence Kummer's test applies, and $\sum_{r=0}^{n} u(r)$ converges, since $\{n \cdot u(n)/u(n+1)\} - (n+1) \geqq k - 1 > 0$.

(ii) If $u(n)/u(n+1) \leqq 1 + 1/n + 1/n \sum_{r=1}^{n} (1/r)$, then $\sum_{r=1}^{n} u(r)$ diverges; we prove first that $\sum_{r=1}^{n} \{1/r \sum_{p=1}^{r} (1/p)\}$ diverges. Since $\sum_{r=1}^{2^n} (1/r) < n + 1$, therefore $\sum_{r=1}^{n} \{1/\sum_{p=1}^{2^r} (1/p)\}$ diverges by comparison with $\sum_{r=0}^{n} (1/(r+1))$, and so, by Cauchy's condensation test, $\sum_{r=1}^{n} \{1/r(\sum_{p=1}^{r} (1/p))\}$ diverges. Writing $v(r) = 1/r(\sum_{p=1}^{r} 1/p)$, we find

$$v(n)/v(n+1) = \{1 + (n+1) \sum_{r=1}^{n} (1/r)\}/n \sum_{r=1}^{n} (1/r)$$
$$= 1 + 1/n + 1/n \sum_{r=1}^{n} (1/r) \geqq u(n)/u(n+1),$$

and therefore $\sum_{r=1}^{n} u(r)$ diverges.

Before summarising the foregoing tests in the Gauss test we must prove the theorem:

If $\sum_{r=1}^{n} u(r)$ *is convergent and* $n\,u(n)$ *steadily decreases to zero, then* $\{\sum_{r=1}^{n} 1/r\} nu(n) \to 0$.

If $n \geqq r$, $nu(n)/r = \{n\,u(n)/r\,u(r)\} u(r) \leqq u(r)$, and so, since $\sum_{r=0}^{n} u(r)$

converges we can choose $n(k)$ so that $\{\sum_{r=n(k)+1}^{n} 1/r\}nu(n) < 1/2k$, for $n > n(k)$. Moreover since $nu(n) \rightarrow 0$, we can determine $r(k)$ such that $\{\sum_{r=1}^{n(k)} 1/r\}nu(n) < 1/2k$, for $n \geqq r(k)$, whence $\{\sum_{r=1}^{n} 1/r\}nu(n) < 1/k$, $n \geqq r(k), n > n(k)$.

The Gauss test. If for some constants α, β, $u(n)/u(n+1) = \alpha + \beta/n + \theta(n)/n^2$, where $|\theta(n)| \leqq M$, then by the ratio test, or Raabe's test, $\alpha > 1$ or $\alpha = 1$, $\beta > 1$ suffice for the convergence of $\sum_{r=0}^{n} u(r)$. Since $\{\sum_{r=1}^{n} 1/r\}(1/n) \rightarrow 0$ by the foregoing theorem, therefore we can find $N(r)$ such that $|\theta(n)|/n^2 < 1/n \sum_{r=1}^{n} (1/r)$, for $n \geqq N(M)$, and so $\alpha < 1$, or $\alpha = 1$ and $\beta \leqq 1$ suffice for the divergence of $\sum_{r=0}^{n} u(r)$.

This completes the preliminary survey and we come now to relative convergence.

Definitions.

The recursive function $s(m, n)$ converges in n, relative to m, if we can determine recursive functions $N(k)$, $M(k, n, n^*)$ such that

$$|s(m, n^*) - s(m, n)| < 1/k, \text{ for all } k \geqq 1,\ n^* \geqq n \geqq N(k), m \geqq M(k, n, n^*).$$

$s(m, n)$ diverges in n, relative to m, if we can determine recursive functions $\lambda(n)$, $M(n, n^*)$ and a constant $k \geqq 1$ such that

$$|s(m, n + \lambda(n)) - s(m, n)| \geqq 1/k, \text{ for } m \geqq M(n, \lambda(n)).$$

$s(m, n) \rightarrow 0$ in n, relative to m, if we determine $N(k)$, $M(k, n)$ such that $|s(m, n)| < 1/k$, for all $k \geqq 1$, $n \geqq N(k)$, $m \geqq M(k, n)$.

Throughout the following theorems $a(m, n)$, $b(m, n)$, $c(m, n)$ are positive, non-zero, recursive functions.

THEOREM 1. *If for some k, $a(m, n) \leqq k < 1$, for $m \geqq M(n)$, then $\sum_{r=0}^{n} \{a(m, r)\}^r$ is convergent in n, relative to m. For $\sum_{r=n}^{N} a(m, r)^r \leqq \sum_{r=n}^{N} k^r$, $m \geqq \max\{M(r), n \leqq r \leqq N\}$.*

THEOREM 1. 1. *If $a(m, n) \geqq 1$, for $n = i(r)$, where $i(r+1) \geqq r + i(r)$,* and for $m \geqq M(n)$ *then $\sum_{r=0}^{n} \{a(m, r)\}^r$ diverges relative to m*; (proof trivial).

THEOREM 2. *If for some $\lambda > 0$, $a(m, n) \leqq \lambda b(m, n)$, for $m \geqq \theta(n)$,* *then the relative convergence of* $\sum_{r=0}^{n} b(m,r)$ *entails that of* $\sum_{r=0}^{n} a(m,r)$, *and the relative divergence of the latter entails that of the former.* For $\sum_{r=m}^{m+n} a(m,r) \leqq \lambda \sum_{r=m}^{m+n} b(m,r)$, $m \geqq \max\{\theta(r), m \leqq r \leqq m+n\}$.

THEOREM 2. 1. *If there are constants* h, H *such that* $0 < h \leqq a(m,0) \leqq H$, $h \leqq b(m,0) \leqq H$ *and* $a(m,n+1)/a(m,n) \leqq b(m,n+1)/b(m,n)$, *for* $m \geqq \theta(n)$, *then the relative convergence of* $\sum_{r=0}^{n} b(m,r)$ *entails that of* $\sum_{r=0}^{n} a(m,r)$, *and the relative divergence of the latter entails that of the former. For* $a(m,n) \leqq (H/h)b(m,n)$, $m \geqq \max\{\theta(r), 0 \leqq r \leqq n-1\}$.

THEOREM 3. $a(m,n+1)/a(m,n) \leqq k < 1$, *for* $m \geqq \psi(n)$, *is a sufficient condition for relative convergence, since* $a(m,n+1)/a(m,n) \leqq k^{n+1}/k^{n}$, *for* $m \geqq \psi(n)$. $a(m,n+1)/a(m,n) \geqq 1$, *for* $m \geqq \phi(n)$, *suffices for relative divergence, since* $\sum_{r=0}^{n} 1$ *diverges.*

THEOREM 4. *If there is a constant* $\beta > 0$ *such that* $c(m,n)\{a(m,n)/a(m,n+1)\} - c(m,n+1) \geqq \beta$, *for* $m \geqq \Gamma(n)$, *and if* $c(m,n)a(m,n) \rightarrow 0$ *in* n, *relative to* m, *then* $\sum_{r=0}^{n} a(m,r)$ *converges relative to* m.

For $\beta \sum_{r=n+1}^{N} a(m,r) \leqq c(m,n)a(m,n)$, $m \geqq \max\,(\Gamma(r,\beta),\ n \leqq r \leqq N-1)$.

THEOREM 5. *If there is a constant* $p > 1$ *such that* $a(m,n)/a(m,n+1) \geqq 1 + p/n$, *for* $m \geqq M(n)$ *then* $\sum_{r=0}^{n} a(m,r)$ *converges relative to* m.

We readily prove $n \cdot a(m,n) \rightarrow 0$ in n, relative to m, and the result then follows from the previous theorem.

THEOREM 5. 1. *If* $a(m,n)/a(m,n+1) \leqq 1 + 1/n + 1/n \sum_{r=1}^{n} (1/r)$, *for* $m \geqq M(n)$, *then* $\sum_{r=0}^{n} a(m,r)$ *diverges relative to* m. (Proof the same as for ordinary divergence.)

THEOREM 6. *If* $f(m,n)$ *is monotonic decreasing in* n, *relative to* m, *i. e. if there is a recursive* $M(n,N)$ *such that* $f(m,N) < f(m,n)$, *for* $N > n$, $m \geqq M(n,N)$, *then* $\sum_{r=0}^{n} f(m,r)$, $\sum_{r=0}^{n} 2^{r} f(m,2^{r})$ *converge and diverge together, relative to* m. (Proof as for ordinary convergence.)

THEOREM 7. *If* $a(m,n)/a(m,n+1) = \alpha + \beta/n + \theta(m,n)/n^{2}$,

$m \geqq \Phi(n)$, $|\theta(m,n)| < M$, *then* $\alpha > 1$, *or* $\alpha = 1$, $\beta > 1$, *are sufficient conditions for relative convergence, and* $\alpha < 1$, *or* $\alpha = 1$, $\beta \leqq 1$ *are sufficient for relative divergence.* This follows from Theorems 3, 5, 5. 1.

We proceed next to sharpen Theorem 7, proving first a series of preparatory theorems.

THEOREM 8. *If* $a > b$, $nb^{n-1} < (a^n - b^n)/(a-b) < na^{n-1}$, $n \geqq 2$.

Proof. If $x > 1$, $n > 1$, then $\sum_{r=0}^{n} x^r > n+1$ (proof by induction), and therefore $(x^{n+1}-1)/(x-1) > n+1$, whence, taking $x = a/b$,

$$(a^{n+1} - b^{n+1})/(a-b) > (n+1)b^n.$$

Similarly, if $x < 1$, $n > 1$, then $\sum_{r=0}^{n} x^r < (n+1)$ and therefore $(1-x^{n+1})/(1-x) < n+1$, and the second half of the theorem follows by taking $x = b/a$.

THEOREM 8. 1. *If* α *is positive, and* p, q *are positive integers,* $p > q$, *then* $(1+\alpha)^p > \{1 + (p/q)\alpha\}^q$.

This is a simple consequence of a special case of the famous theorem of the means, viz.

$(ma + nb)^{m+n}/(m+n) > a^m b^n$, $a \neq b$, m, n positive, non-zero integers. (There are several algebraic proofs of the theorem of the means which are valid in the present system, for instance Cauchy's proof as it is given in [4]).

Take $m = q$, $m + n = p$, $a = 1 + (p/q)\alpha$, $b = 1$ and the theorem follows. Similarly, if $p < q$, taking $m = p$, $m + n = q$, $b = 1$, $a = 1 + \alpha$, we have $(1+\alpha)^p < \{1 + (p/q)\alpha\}^q$.

THEOREM 9. *If* p, q, x *are integers,* $x > 1$, $q > 1$, $p \geqq 1$, *and if* x_k *is the greatest integer such that* $(x_k)^q \leqq x^p \cdot 2^{kq}$, *then* $x_n/2^n$ *is convergent.*

Proof. Since $(2x_k)^q \leqq x^p \cdot 2^{(k+1)q}$, therefore $x_{k+1} \geqq 2x_k$ and so $x_k/2^k$ is monotonic increasing, and since x_0 is the greatest integer such that $(x_0)^q \leqq x^p$, and $1^q < x^p$, then $x_0 \geqq 1$ and so $x_k > 2^k$. From $(x_k)^q \leqq x^p \cdot 2^{kq} < x^{pq} \cdot 2^{kq}$ it follows that $x_k < x^p \cdot 2^k$.[2] Hence

$$0 \leqq x^p - (x_k/2^k)^q < \{(x_k+1)^q - (x_k)^q\}/2^{kq} < q(x_k+1)^{q-1}/2^{kq}$$

(by Theorem 8)

$$< q\{x^p + 1/2^k\}^{q-1}/2^k \leqq q\{x^p+1\}^{q-1}/2^k \to 0,$$

i. e. $(x_k/2^k)^q$ converges to x^p.

---

[2] Since $t^q - 1 = \{\sum_{r=0}^{q-1} t^r\}(t-1)$, it follows that if $t > 0$, then $t \gtrless 1$ according as $t^q \gtrless 1$.

Writing $x_k/2^k = y_k$, so that y_k is monotonic increasing and $y_k{}^q$ converges, then, since $\{(y_{k+r})^q - (y_k)^q\}/\{y_{k+r} - y_k\} > q(y_k)^{q-1} > q$, i.e. $y_{k+r} - y_k < \{(y_{k+r})^q - (y_k)^q\}/q$, so that y_k converges, which completes the proof.

Definition. We define, for $p \geqq 1$, $q > 1$, $x > 1$, $\chi(x, p, q, k) = x_k/2^k$ (where x_k is given alone) and $\chi(1, q, p, k) = 1$.

THEOREM 10. $\sum_{r=1}^{n} 1/\chi(r, p, q, k)$ *is convergent in* n, *relative to* k, *if* $p > q$ *and divergent in* n, *relative to* k, *if* $p \leqq q$.

Since $\{\chi(x, p, q, k)\}^q \to x^p$, therefore, writing $a_k = \chi(x+1, p, q, k)/\chi(x, p, q, k)$, we have if $p > q$, $(a_k)^q \to (1 + 1/x)^p > (1 + p/qx)^q$ by Theorem 8.1; the convergence of a_k determines a recursive function $Q(x)$ such that $|(1 + 1/x)^p - (a_k)^q| < (1 + 1/x)^p - (1 + p/qx)^q$ for $k \geqq Q(x)$, and therefore $(a_k)^q > (1 + p/qx)^q$, i.e. $a_k > 1 + p/qx$ for $k \geqq Q(x)$. It follows, by Theorem 5, that $\sum_{r=1}^{n} 1/\chi(r, p, q, k)$ converges relative to k, if $p > q$.

If $p = q$, $\chi(n, p, q, k) = n$, and so $\sum_{r=1}^{n} 1/\chi(r, p, q, k)$ diverges. If $p < q$, $a_k < 1 + p/qx$ for $k \geqq Q^*(x)$, whence, by Theorem 5.1, $\sum_{r=1}^{n} 1/\chi(r, p, q, k)$ diverges relative to k.

THEOREM 11. *If* $\phi(n, p, q, k) = n/\chi(n, p, q, k)$ *then, for* $p > q$, $\phi(n, p, q, k)$ *is monotonic decreasing to zero, relative to* k.

For

$$\begin{aligned}\phi(n, p, q, k)/\phi(n+1, p, q, k) &> \{n/(n+1)\}\{1 + p/qn\} \\ &= 1 + ((p/q) - 1)/(n+1)\end{aligned}$$

if $k \geqq Q(n)$, $p > q$, so that $\phi(n, p, q, k)$ is monotonic decreasing, relative to k, and further, if $N > n$,

$$\begin{aligned}\phi(N, p, q, k)/\phi(n, p, q, k) &< 1/\prod_{r=n+1}^{N} \{1 + ((p/q) - 1)/r\} \\ &< 1/((p/q) - 1) \sum_{r=n+1}^{N} (1/r) \to 0,\end{aligned}$$

$$k \geqq \max\{Q(r), n \leqq r \leqq N - 1\},$$

i.e. $\phi(N, p, q, k) \to 0$ relative to k.

THEOREM 11.1. *If* $p > q$, $\{\sum_{r=1}^{n} (1/r)\} n/\chi(n, p, q, k) \to 0$, *relative to* k.

If $n > r$, by Theorem 11,

$$\phi(n, p, q, k)/\phi(r, p, q, k)\chi(r, p, q, k) < 1/\chi(r, p, q, k)$$

for $k \geqq \max\{Q(m), r \leqq m \leqq n-1\}$. Since $\sum_{r=1}^{n} 1/\chi(r, p, q, k)$ converges relative to k, if $p > q$, given $x \geqq 1$, we can determine $n(x)$ and $\delta(n, x)$ such that, $\sum_{r=n(x)}^{n} 1/\chi(r, p, q, k) < 1/2x$, for $n \geqq n(x)$, $k \geqq \delta(n, x)$. Furthermore, since $\phi(n, p, q, k) \to 0$, relative to k, we can determine $\nu(x)$ and $\eta(x)$ such that $\{\sum_{r=1}^{n(x)} (1/r)\}\phi(n, p, q, k) < 1/2x$, for $n \geqq \nu(x)$, $k \geqq \eta(n, x)$. Hence

$$\{\sum_{r=1}^{n} (1/r)\}\phi(n, p, q, k) < 1/x,$$

$n \geqq \max(\nu(x), n(x))$, $k \geqq \max\{\eta(n, x), \delta(n, x), [\max\{Q(m), n(x) \leqq m \leqq n-1\}]\}$ i. e. $\{\sum_{r=1}^{n} (1/r)\}\phi(n, p, q, k) \to 0$ relative to k.

We are now in a position to state the analogue, for relative convergence, of the Gauss test.

THEOREM 12. *If* $a(n, k)/a(n+1, k) = \alpha + \beta/n + \theta(n, k)/\chi(n, p, q, k)$, *for* $p > q$, $|\theta(n, k)| < M$, *and* $k \geqq A(n)$, *then* $\alpha > 1$ *or* $\alpha = 1$, $\beta > 1$ *are sufficient conditions for the convergence of* $\sum_{r=0}^{n} a(r, k)$ *relative to* k, *and* $\alpha < 1$ *or* $\alpha = 1$, $\beta \leqq 1$ *are sufficient conditions for relative divergence.*

We have to consider only the cases $\alpha = 1$, $\beta > 1$; $\alpha = 1$, $\beta \leqq 1$. If $\beta > 1$, since $\theta(n, k)\phi(n, p, q, k) \to 0$ relative to k, we can find N_0, $K(n)$ such that

$$|n \cdot \theta(n, k)/\chi(n, p, q, k)| < (1/2)(\beta - 1), \text{ for } n \geqq N_0,\ k \geqq K(n),$$

and so

$$a(n, k)/a(n+1, k) > 1 + (1/2)(\beta + 1)/n, \text{ for } n \geqq N_0,\ k \geqq \max\{K(n), A(n)\},$$

which suffices for relative convergence, by Theorem 5.

Similarly, if $\beta < 1$, $a(n, k)/a(n+1, k) < 1 + (1/2)(\beta + 1)/n$, which proves relative divergence, by Theorem 5. 1.

If $\beta = 1$, since $\{\sum_{r=1}^{n} (1/r)\}\phi(n, p, q, k) \to 0$, relative to k, therefore we can find n_0 and $J(n)$ such that $|\theta(n, k)/\chi(n, p, q, k)| < 1/n \sum_{r=1}^{n} (1/r)$, for $n \geqq n_0$, $k \geqq J(n)$, whence $\sum_{r=0}^{n} a(r, k)$ is divergent, relative to k, by Theorem 5. 1.

Appendix.

The index laws for $\chi(n, p, q, k)$.

1. If $r > 0$, $\chi(n, pr, qr, k) = \chi(n, p, q, k)$.

Writing $n_k = 2^k \cdot \chi(n, p, q, k)$, then $n_k^{qr} \leqq n^{pr} \cdot 2^{kqr}$ and $(n_k + 1)^{qr} > n^{pr} \cdot 2^{kqr}$ so that n_k is the greatest integer such that $n_k^{qr} \leqq n^{pr} \cdot 2^{kqr}$, whence the result follows.

Accordingly we may write $\chi(n, p, q, k)$ *in the form* $\chi(n, p/q, k)$.

2. $\chi(ab, p/q, k) - \chi(a, p/q, k) \cdot \chi(b, p/q, k) \rightarrow 0$ in k.

We have $\{\chi(a, p/q, k)\}^q \leqq a^p$, $\{\chi(b, p/q, k)\}^q \leqq b^p$, and therefore

$$\{2^k\chi(a, p/q, k) \cdot 2^k\chi(b, p/q, k)\}^q \leqq (ab)^p 2^{2kq};$$

but $2^k\chi(ab, p/q, k) + 1$ is the least integer such that

$$\{2^k\chi(ab, p/q, k) + 1\}^q > (ab)^p \cdot 2^{kq},$$

and therefore

$$2^{2k}\chi(ab, p/q, k) + 2^k > 2^k\chi(a, p/q, k) \cdot 2^k\chi(b, p/q, k),$$

i. e.

$$\chi(ab, p/q, k) - \chi(a, p/q, k) \cdot \chi(b, p/q, k) > -1/2^k.$$

Similarly

$$(\chi(a, p/q, k) + 1/2^k)(\chi(b, p/q, k) + 1/2^k) > \chi(ab, p/q, k),$$

whence

$$\begin{aligned} -1/2^k &< \chi(ab, p/q, k) - \chi(a, p/q, k) \cdot \chi(b, p/q, k) \\ &< \{\chi(a, p/q, k) + \chi(b, p/q, k) + 1/2^k\}/2^k, \end{aligned}$$

which proves the theorem, since $\chi(a, p/q, k)$, $\chi(b, p/q, k)$ converge in k.

3. If $p \geqq 1$, $q > 1$, $r \geqq 1$, $s > 1$, $n \geqq 1$, then

$$\chi(n, p/q, k) \cdot \chi(n, r/s, k) - \chi(n, p/q + r/s, k) \rightarrow 0 \text{ in } k.$$

(Proof the same as in **2** above.)

Definition. We define, for $a \geqq 1$, $b > 1$,

$$\chi_f(a, b, p/q, k) = \chi(a, p/q, k)/\chi(b, p/q, k).$$

Since $\chi(a, p/q, k)$, $\chi(b, p/q, k)$ are both convergent in k, and $\chi(b, p/q, k) > 1$, therefore $\chi_f(a, b, p/q, k)$ is convergent in k.

4. If $c \geqq 1$, $\chi_f(ac, bc, p/q, k) - \chi_f(a, b, p/q, k) \rightarrow 0$ in k.

For $\chi_f(ac, bc, p/q, k) - \chi_f(a, b, p/q, k)$

$$= \frac{\chi(ac, p/q, k)}{\chi(bc, p/q, k)} - \frac{\chi(a, p/q, k)}{\chi(b, p/q, k)} \cdot \frac{\chi(c, p/q, k)}{\chi(c, p/q, k)}$$

$\rightarrow 0$, in k, by **2**. *Accordingly we may write* $\chi_f(a/b, p/q, k)$ *for* $\chi_f(a, b, p/q, k)$.

5. $\chi_f(\chi(x, p/q, k), r/s, \lambda)$ is convergent in k relative to λ and

$$\chi(x, pr/qs, k) - \chi_f(\chi(x, p/q, k), r/s, \lambda) \rightarrow 0$$

in k, relative to λ.

Proof. Write $x_k = 2^k \cdot \chi(x, p/q, k)$, so that

$$\chi_f(\chi(x, p/q, k), r/s, \lambda) = \chi(x_k, r/s, \lambda)/\chi(2^k, r/s, \lambda).$$

Keeping k fixed, $\{\chi(x_k, r/s, \lambda)\}^s \rightarrow (x_k)^r$ and $\{\chi(2^k, r/s, \lambda)\}^s \rightarrow 2^{kr}$, and, further, $(x_k/2^k)^q \rightarrow x^p$, in k, wherefore we can find $K(n)$ and $L(n, k)$ such that

$$0 \leqq x^{pr} - (x_k/2^k)^{qr} < 1/n, \text{ for } k \geqq K(n),$$

$$|\{x_k/2^k\}^{qr} - \{\chi(x_k, r/s, \lambda)/\chi(2^k, r/s, \lambda)\}^{qs}| < 1/n, \text{ for } \lambda \geqq L(n, k),$$

and so

$$| x^{pr} - \{\chi(x_k, r/s, \lambda)/\chi(2^k, r/s, \lambda)\}^{qs} | < 2/n, \text{ for } k \geqq K(n),\ \lambda \geqq L(n, k),$$

i. e. $\{\chi_f(\chi(x, p/q, k), r/s, \lambda)\}^{qs}$ is convergent to x^{pr} in k, relative to λ; denoting $\chi_f(\chi(x, p/q, k), r/s, \lambda)$ by $\Omega(k, \lambda)$ we see that, since $x_k > 2^k$ and $\chi(y, r/s, \lambda)$ increases with y, therefore $\Omega(k, \lambda) \geqq 1$, and so for any n, N, by Theorem 8, if $p \geqq 1$, $| \Omega(N, \lambda) - \Omega(n, \lambda)| \leqq | \Omega(N, \lambda)^p - \Omega(n, \lambda)^p | /p$, whence it follows that $\Omega(k, \lambda)$ converges in k relative to λ. Furthermore $\{\chi(x, pr/qs, k)\}^{qs} \rightarrow x^{pr}$ and so

$$\{\chi(x, pr/qs, k)\}^{qs} - \{\Omega(k, \lambda)\}^{qs} \rightarrow 0$$

in k, relative to λ; since $\Omega(k, \lambda) \geqq 1$ and $\chi(x, pr/qs, k) \geqq 1$, therefore

$$| \chi(x, pr/qs, k) - \Omega(k, \lambda) | \leqq |\{\chi(x, pr/qs, k)\}^{qs} - \{\Omega(k, \lambda)\}^{qs} | /qs$$

(if $q \geqq 1$, $s \geqq 1$) and so $\chi(x, pr/qs, k) - \Omega(k, \lambda) \rightarrow 0$ in k, relative to λ.

UNIVERSITY COLLEGE, LEICESTER, ENGLAND.

BIBLOGRAPHY.

1. H. B. Curry, "Formalisation of recursive arithmetic," *American Journal of Mathematics*, vol. 63 (1941), pp. 263-282.
2. R. L. Goodstein, "Function theory in an axiom-free equation calculus," *Proceedings of the London Mathematical Society* (2), vol. 48 (1945), pp. 401-434.
3. D. Hilbert and P. Bernays, *Grundlagen der Mathematik*, Berlin, I (1934), II (1939).
4. G. H. Hardy, J. E. Littlewood, G. Pólya, *Inequalities*, Cambridge (1934).

JACOBIAN EXTENSORS.*

By Homer V. Craig and William T. Guy, Jr.

---

1. Introduction. The purpose of the present paper is to extend the general notions of *contravariance* and *covariance* so as to include what we shall call *Jacobian-contravariance* and *Jacobian-covariance* or briefly *j-contravariance* and *j-covariance.* As the title perhaps suggests, the essence of this innovation is the introduction of quantities which follow a transformation law which is formally equivalent to that of tensor analysis but differs in content by the presence of certain coefficients $X^{a}{}_{\rho}$ which depend upon the Jacobian and are analogous to the extensor coefficients X sup αa inf ρr. To be more specific, the new transformation equations contain, in addition to the coefficients $X^{a}{}_{r}$ $(X^{a}{}_{r} = \partial x^{a}/\partial \bar{x}^{r})$ and X sup αa inf ρr of tensor and extensor analysis, certain coefficients $X^{a}{}_{\rho}$ which are related to the weighted Jacobian $\underline{X}$ $(\underline{X} = x/\bar{x})^{w})$ in the same way that the coefficients X sup αa inf ρr are related to the quantities $X^{a}{}_{r}$. The symbol X sup αa inf ρr denotes the partial derivative of $x^{\alpha a}$ with respect to $\bar{x}^{\rho r}$ and the relationship just mentioned is

$$X^{\alpha a}{}_{\rho r} = \binom{\mathrm{A}}{\mathrm{P}} X^{a}{}_{r}{}^{(\mathrm{A}-\mathrm{P})}.$$

Here $\mathrm{A} = \alpha$, $\mathrm{P} = \rho$ and the enclosed superscripts indicate differentiation with respect to the curve parameter t. The capital indices are introduced to forestall summation.

We shall find that Jacobian extensors include weighted and absolute tensors and scalars as special cases and that their properties are so similar to those of ordinary extensors that much of the older theory may be taken over without proof. In addition, we shall show that given an ordinary connection one may develop readily its Jacobian counterpart, which we call a *Jacobian connection.* Finally, the Jacobian connections are such that they enable us to express the higher order intrinsic derivatives of weighted tensors as contractions of certain extensors. This provides a very satisfying theory of the algebraic structure of these rather complicated derived tensors.

The material to be presented falls into two rather sharply delineated subdivisions. Part A is concerned entirely with developments which are valid in any N-dimensional space which bears a coordinate system, while

---

* Received December 1, 1948. Presented to the Society, November 27, 1948.

Part B presumes a space with symmetric connection. The symmetry of the connection facilitates the work by providing for the use of geodesic coordinates.

Part A.

In the present major division, we assume that there is given an N-dimensional space, a coordinate system, and the group of one-to-one differentiable coordinate transformations.

2. Notation. The notation employed is for the most part that of certain related works, particularly numbers [3], [4], [5] of the appended bibliography. Accordingly, we describe it rather briefly. Enclosed Greek and numerical indices as well as primes indicate differentiation with respect to a curve parameter t. Repeated lower case Latin and Greek indices engender summations from 1 to N and from 0 to M, respectively, unless the contrary is indicated; while repeated capital indices do *not* generate sums. Further, indices at the last of the alphabet such as r, s, t, u, v, w, ρ, σ, τ indicate that the symbol of which they are a part belongs to coordinate system $(\bar{x})$. Thus we write x^r for $\bar{x}^r$. Indices at the first of the alphabet a, b, c, d, e, α, β, γ, δ, ϵ are correlated to system (x). Finally, the symbols $(x/\bar{x})$, $(\bar{x}/x)$ denote, respectively, the Jacobians associated with the coordinate transformations $x^a = x^a(\bar{x})$, $\bar{x}^r = \bar{x}^r(x)$, while $\underline{X}$ and $\bar{X}$ stand for these Jacobians raised to the w power with w a given positive integer or zero, thus $\underline{X} = (x/\bar{x})^w$.

Special convention. All terms which contain symbols bearing numerical indices out of their natural range are to be given the value zero. For example $X^{(\alpha-\rho)} = 0$ whenever ρ exceeds α, for we take the range of indices which indicate the order of differentiation to be zero and the positive integers instead of extending this range by a definition.

3. The Jacobian symbols $X^{\alpha}{}_{\rho}$, $X^{\rho}{}_{\alpha}$. The coefficients of the components in the transformation equation for Jacobian extensors are defined as follows:

Definition 3. 1. $X^{\alpha}{}_{\rho} = \binom{\mathrm{A}}{\mathrm{P}} \underline{X}^{(\mathrm{A}-\mathrm{P})}$ $\qquad X^{\rho}{}_{\alpha} = \binom{\mathrm{P}}{\mathrm{A}} \bar{X}^{(\mathrm{P}-\mathrm{A})}$

$\mathrm{A} = \alpha$, $\mathrm{P} = \rho$.

With regard to this definition, it should be borne in mind that the new symbols vanish, by virtue of our special convention, whenever the subscript exceeds the superscript. A second and very important property of these symbols is expressed by

THEOREM 3.1. *If $\bar{\bar{x}}$, x, and $\bar{x}$ are any three coordinate systems of the infinite group of tensor analysis and if we correlate the indices λ, α, and ρ to these coordinate systems, then*

$$X^\lambda{}_\alpha X^\alpha{}_\rho = X^\lambda{}_\rho.$$

Proof. By virtue of Definition 3.1 and the fact that the *effective* range of α is from ρ to λ since an X vanishes whenever its subscript exceeds its superscript, we may transcribe the left member as follows

$$X^\lambda{}_\alpha X^\alpha{}_\rho = \sum_{\alpha=\mathrm{P}}^{\Lambda} \binom{\Lambda}{\alpha}\binom{\alpha}{\mathrm{P}} \bar{X}^{(\lambda-\alpha)} \underline{X}^{(\alpha-\mathrm{P})}.$$

We now replace the dummy index α in the right member with $\Lambda - \beta$ with β ranging from 0 to $\Lambda - \mathrm{P}$. The result is

$$\sum_{\beta=0}^{\Lambda-\mathrm{P}} \binom{\Lambda}{\Lambda-\beta}\binom{\Lambda-\beta}{\mathrm{P}} \bar{X}^{(\beta)} \underline{X}^{(\Lambda-\mathrm{P}-\beta)}$$

which upon substituting $\binom{\Lambda}{\mathrm{P}}\binom{\Lambda-\mathrm{P}}{\mathrm{B}}$ for $\binom{\Lambda}{\Lambda-\mathrm{B}}\binom{\Lambda-\mathrm{B}}{\mathrm{P}}$ becomes

$$\binom{\Lambda}{\mathrm{P}} \sum_{\beta=0}^{\Lambda-\mathrm{P}} \binom{\Lambda-\mathrm{P}}{\beta} \bar{X}^{(\beta)} \underline{X}^{(\Lambda-\mathrm{P}-\beta)}.$$

But by the Leibniz rule for differentiating a product and the fact that $\bar{X}\underline{X} = (\bar{\bar{x}}/x)^w \quad (x/\bar{x})^w = (\bar{\bar{x}}/\bar{x})^w = \bar{\underline{X}}$, we may transform this last expression successively into the members of the equality

$$\binom{\Lambda}{\mathrm{P}} (\bar{X}\underline{X})^{(\Lambda-\mathrm{P})} = \binom{\Lambda}{\mathrm{P}} \bar{\underline{X}}^{(\Lambda-\mathrm{P})} = X^\lambda{}_\rho.$$

Thus we have proved that $X^\lambda{}_\alpha X^\alpha{}_\rho = X^\lambda{}_\rho$ and the theorem is established.

An immediate consequence of this theorem is the

COROLLARY. $X^\rho{}_\alpha X^\alpha{}_\sigma = \delta^\rho{}_\sigma.$

Proof. $X^\rho{}_\sigma = \binom{\mathrm{P}}{\Sigma} 1^{(\mathrm{P}-\Sigma)} = \delta^\rho{}_\sigma.$

4. Jacobian extensors. As we have indicated in the introduction, the essential point of distinction between ordinary and Jacobian extensors is the presence of the symbols $X^\alpha{}_\rho$ in the transformation equations of the latter quantities. To illustrate, the transformation equations for the two types of Jacobian extensors of the first order of weight w $(w > 0)$ and range M are as follows

(4.1) $U^\alpha = U^\rho X^\alpha{}_\rho$; (4.2) $V_\alpha = V_\rho X^\rho{}_\alpha.$

The first of these quantities U will be described by saying that it is *j-contravariant,* while the quantities such as V, which follow the transformation equation (4.2), will be said to be *j-covariant.* Similarly, the single Greek indices such as the α on U^α and the α on V_α will be said to be *j-contravariant* and *j-covariant indices,* respectively. The distinction between these two equations, it will be observed, hinges on the fact that w by convention is never negative. The concept extensor when enlarged so as to include the present innovation, may be formulated as follows:

DEFINITION 4.2. *Let there be given at a point P of a parameterized arc of class c^M one set of labelled numbers $E^{\alpha.\gamma c.e}_{\beta.\delta d.f}$ for each coordinate system of the infinite collection of tensor analysis. Further, let x and $\bar{x}$ denote any two coordinate systems of this group and let $E^{\alpha.\gamma c.e}_{\beta.\delta d.f}$ and $E^{\rho.\tau t.u}_{\sigma.\omega w.v}$ be associated with systems x and $\bar{x}$, respectively. If these quantities are related according to the equation*

$$E^{\alpha.\gamma c.e}_{\beta.\delta d.f} = E^{\rho.\tau t.u}_{\sigma.\omega w.v} X_\rho{}^\alpha X_\beta{}^\sigma X_{\tau t}{}^{\gamma c} X_{\delta d}{}^{\omega w} X_u{}^e X_f{}^v \tag{4.3}$$

then, we shall say that they are the components of an extensor of range M and weight w which is j-contravariant, j-covariant, excontravariant, excovariant, contravariant, and covariant—each of order one.

The extension of this definition to higher orders of contravariance and covariance and the generalization to the case in which not all of the Greek indices have the same range are obvious and do not require special enunciation. It should be observed in passing, however, that if a Jacobian superscript is confined to the value zero or a Jacobian subscript to the value M (in the case of doublet indices these are tensor values), then the effect of these indices is to weight the extensor in the ordinary sense of the term; in other words, the effect is to introduce a weighted Jacobian as a multiplier. To illustrate, if α in $E_f{}^\alpha$ is confined to the value zero, then the transformation equation

$$E_f{}^\alpha = E_v{}^\rho X_\rho{}^\alpha X_f{}^v \text{ reduces to } E_f{}^0 = (\bar{x}/x)^{-w} E_v{}^0 X_f{}^v;$$

while the introduction of the restriction $\beta = M$ into

$$E_\beta{}^e = E_\sigma{}^u X_\beta{}^\sigma X_u{}^e \text{ yields } E_M{}^e = (\bar{x}/x)^w E_M{}^u X_u{}^e.$$

Thus $E_f{}^0$ is a tensor of weight $-w$, while $E_M{}^e$ is a tensor of weight w, and we can see that *the concept Jacobian extensor encompasses that of weighted tensor as a special case.*

Reduced range. Obviously, the property just noted admits of an immediate extension since it is true in general that $X_\rho{}^\alpha$ vanishes whenever ρ exceeds α. Thus, in particular, we see from the transformation equation for $E_\beta{}^\alpha$, namely, $E_\beta{}^\alpha = E_\sigma{}^\rho X_\rho{}^\alpha X_\beta{}^\sigma$, that the components of $E_\beta{}^\alpha$ which have $\alpha \leqq \theta$ and $\beta \geqq \Phi$ are independent of the components in the other coordinate systems which do not meet the corresponding restrictions.

Transitive property. Since $X_\rho{}^\alpha X_\beta{}^\rho = \delta_\beta{}^\alpha$ and $X_\alpha{}^\rho X_\lambda{}^\alpha = X_\lambda{}^\rho$, it follows at once that the inverse of Equation 4. 3 is of the same type as the original and that the transformation law for Jacobian extensors is transitive.

Invariance, sums, products, and contractions. From the fact that 4. 3 is linear and homogeneous it follows as in tensor analysis that extensor equations of the Jacobian type are invariant equations, and that the familiar rules regarding the sums and products of tensors have their counterparts in the new theory. Similarly, as in extensor analysis, there are $M+1$ contractions of a mixed second order Jacobian extensor. The form of these contractions is given by the following statement:

THEOREM 4. 1. *If $E_\beta{}^\alpha$ is a Jacobian extensor of the type indicated by the indices, then for each fixed value of θ from 0 to M, inclusive, $\sum\limits_{\alpha=\theta}^{M}\binom{\alpha}{\theta}E^{\alpha-\theta}{}_\alpha$ is an absolute scalar or invariant.*

Proof. Applying successively the transformation equation for Jacobian extensors, Definition 3. 1, and the formula

$$\binom{\alpha}{\theta}\binom{\alpha-\theta}{\rho-\theta}\binom{\alpha}{\rho}^{-1}=\binom{\rho}{\theta},$$

we get the chain of equalities

$$\sum_{\alpha=0}^{M}\binom{\alpha}{\theta}E^{\alpha-\theta}{}_\alpha=\sum_{\alpha,\rho=\theta}^{M}\binom{\alpha}{\theta}E^{\rho-\theta}{}_\sigma X^{\alpha-\theta}{}_{\rho-\theta}X^\sigma{}_\alpha$$

$$=\sum_{\alpha,\rho=\theta}^{M}\binom{\alpha}{\theta}\binom{\alpha-\theta}{\rho-\theta}\binom{\alpha}{\rho}^{-1}E^{\rho-\theta}{}_\sigma X^\alpha{}_\rho=\sum_{\rho=\theta}^{M}\binom{\rho}{\theta}E^{\rho-\theta}{}_\sigma\delta^\sigma{}_\rho=\sum_{\rho=\theta}^{M}\binom{\rho}{\theta}E^{\rho-\theta}{}_\rho, \quad \alpha\geqq\rho,$$

and the theorem is established.

Remark. This proof is obviously sufficiently typical to enable us to infer the corresponding theorem for higher order extensors. The additional indices whether single Greek, doublet, or single Latin would call for only trivial modification.

Remark. This contraction process, except for the cases $\theta=0$, $\theta=M$, is more complicated than that of tensor analysis. However, it will be found to be easy to reproduce if we note that the essential point is the pattern of α and θ in the expression $\sum\limits_{\alpha=\theta}^{M}\binom{\alpha}{\theta}E^{\alpha-\theta}{}_{\alpha}$. Strictly as a memory device, this pattern may be regarded as generated by rotating the $\alpha\theta$ in $E^{\alpha-\theta}$ first clockwise into the binomial coefficient $\binom{\alpha}{\theta}$ and then counterclockwise to the summation symbol, thus $\sum\limits_{\alpha=\theta}$. With this in mind, we denote $\sum\limits_{\alpha=\theta}^{M}\binom{\alpha}{\theta}E^{\alpha-\theta}{}_{\alpha}$ by the abridged symbol $RE^{\alpha-\theta}{}_{\alpha}$. In this notation the contraction theorem asserts that $RE^{\alpha-\theta\cdots}{}_{\alpha\cdots}$ is an extensor of the type indicated by the remaining indices.

Quotient law. Following the pattern of the parent theories, there is associated with the contraction theorem a quotient law involving a partly arbitrary Jacobian extensor A. The degree of arbitrariness is of course that required in the proof. The theorem in question is as follows:

THEOREM 4.2. *If $E^{\cdots}{}_{\cdots}$ is independent of A, and if for some admissible value of θ, $RE^{\alpha-\theta\cdots}{}_{\cdots}A_{\alpha}$ is an extensor of the type indicated by the free indices for arbitrary choice of the Jacobian extensor A_{α}, then the same may be asserted of $E^{\alpha\cdots}{}_{\cdots}$ for α in the reduced range 0 to $M-\theta$.*

Proof. As in the case of tensors and ordinary extensors the proof consists of merely writing out the given transformation equation, "factoring out" A_{α} and then assigning A_{α} certain special values. The details are as follows. We have given that for some fixed value of θ

$$RE^{\alpha-\theta\cdots}{}_{\cdots}A_{\alpha}-RE^{\rho-\theta\cdots}A_{\rho}X=0,$$

if X represents the set of Jacobian symbols and partial derivatives that are tied into the missing indices on $\bar{E}$. Replacing A_{ρ} with $A_{\alpha}X^{\alpha}{}_{\rho}$ (the range of α here may be taken to be θ to M since $\rho\geqq\theta$), we get

$$[RE^{\alpha-\theta\cdots}{}_{\cdots}-RE^{\rho-\theta\cdots}{}_{\cdots}X^{\alpha}{}_{\rho}]A_{\alpha}=0.$$

We now assign A_{α} the special values $\delta^{\phi+\theta}{}_{\alpha}$, $\Phi+\theta\leqq M$, and simultaneously replace the dummy index ρ with $\rho+\theta$, with the new ρ ranging from 0 to $M-\theta$. The result is

$$\binom{\Phi+\theta}{\theta}E^{\Phi\cdots}{}_{\cdots}-\sum_{\rho=0}^{M-\theta}\binom{\rho+\theta}{\theta}E^{\rho\cdots}{}_{\cdots}X^{\phi+\theta}{}_{\rho+\theta}=0.$$

An easy calculation will show that $\binom{\mathrm{P}+\theta}{\theta}\binom{\Phi+\theta}{\mathrm{P}+\theta}\binom{\Phi}{\mathrm{P}}^{-1}=\binom{\Phi+\theta}{\theta}$.

Hence $\binom{\mathrm{P}+\theta}{\theta}X^{\Phi+\theta}{}_{\mathrm{P}+\theta}=\binom{\Phi+\theta}{\theta}X^{\phi}{}_{\rho}$ and the Theorem follows.

Remark. For $M - B$ any number from 0 to M, A_a will reduce to $\delta^{M-B}{}_{a}$ at $t = c$, if $B!A_a = [(t-c)^B]^{(M-a)}$. Later we shall encounter Jacobian extensors satisfying the relationship $A_a = \binom{M}{\mathrm{A}} A_{M}{}^{\cdot(M-\mathrm{A})}$. Thus the foregoing proof will be valid if it is required, in addition, that $A_a = \binom{M}{\mathrm{A}} A_{M}{}^{\cdot(M-\mathrm{A})}$ in all coordinate systems.

Remark. The corresponding proposition for A^a is

THEOREM 4. 3. *If $E^{\cdots}{}_{\ldots}$ is independent of the arbitrary extensor A and for some admissible value of θ, $RA^{a-\theta}E^{\cdots}{}_{a\ldots}$ is an extensor of the type indicated by the free indices, then the same may be asserted of $E^{\cdots}{}_{a\ldots}$ for α ranging from θ to M.*

The proof of this theorem may be obtained from the proof of the corresponding theorem for ordinary extensors, making the obvious changes. See [3],[1] page 272.

Examples of Jacobian extensors. Obviously, as we have tacitly assumed in the preceding proof, Jacobian extensors may be manufactured by arbitrarily assigning values to the components in any one coordinate system and then defining the components in the other coordinate systems by means of the transformation law itself. In addition, we may, of course, take over without proof many of the theorems of ordinary extensor analysis because of the formal similarity between absolute and Jacobian extensors. Certain of these theorems will point the way to the construction of Jacobian extensors. As cases in point, we may cite the following propositions:

THEOREM 4. 4. *If V^0 is a completely reduced Jacobian extensor and hence of range* 0 *(or, in other words, if V^0 is a scalar of weight $-W$), then the quantities $V^{0(a)}$, $0 \leqq \alpha \leqq M$, constitute a Jacobian extensor of the range* 0 *to M.*

THEOREM 4. 5. *If V_M is a completely reduced Jacobian extensor of the type indicated, i. e., a scalar of weight W, then the quantities V_a defined by $V_a = \binom{M}{\mathrm{A}} V_{M}{}^{\cdot(M-\mathrm{A})}$, $\mathrm{A} = \alpha$, $0 \leqq \alpha \leqq M$ are the components of a Jacobian extensor of range* 0 *to M.*

Remark. Since the proofs of these theorems are very short and essentially the same, it will perhaps be defensible to present one of them in place of reexamining the corresponding proofs for ordinary extensors. Accordingly, we give the argument for Theorem 4. 4 in outline form.

---

[1] Numbers in brackets refer to references in the appended bibliography.

Proof. From the transformation equation $\bar{V}^{\rho} = V^{a}X^{\rho}{}_{a}$, we derive as the transformation equation for the completely reduced Jacobian extensor, the relationship $\bar{V}^{0} = V^{a}X^{0}{}_{a} = V^{0}\bar{X}$. Differentiating ρ times by the Leibnitz rule, we get

$$\bar{V}^{0(\rho)} = (V^{0}\bar{X})^{(\rho)} = \sum_{a=0}^{\mathrm{P}} V^{0(a)} \binom{\mathrm{P}}{\alpha} \bar{X}^{(\mathrm{P}-a)} = V^{0(a)}X^{\rho}{}_{a}.$$

These theorems provide the means for expanding the range of completely reduced Jacobian extensors (weighted scalars) to the full range $(0, M)$. A moment's consideration will show that similar theorems exist for higher order weighted tensors. To illustrate, the quantities $E^{\alpha a.\beta b.\gamma c}_{\delta d}$ defined by $E^{\alpha a.\beta b.\gamma c}_{\delta d} = \left\{\begin{matrix}\mathrm{A}, \mathrm{B}, \Gamma \\ M, \Delta\end{matrix}\right\} T^{abc}{}_{d}{}^{(\mathrm{A}+\mathrm{B}+\Gamma-2M-\Delta)}$ constitute the components of an absolute extensor provided $T^{abc}{}_{d}$ is an absolute tensor.[2] Furthermore, the conversion of the fixed Jacobian subscript M in Theorem 4. 5 into a full range Jacobian subscript by differentiation with respect to the parameter is, as we have pointed out, perfectly analogous to the conversion of a tensor subscript d into the doublet subscript δd in the illustration just cited. Similarly, the conversion of a Jacobian superscript zero (this indicates weight $-w$ in the usual nomenclature) into a full range Jacobian superscript α is formally exactly equivalent to the conversion of a tensor superscript a into the doublet superscript αa. Accordingly, we may take over without proof the following theorems from [5], pp. 335, 336.

THEOREM 4. 6. *If* $T^{0.bc}$ *is a second order contravariant tensor of weight* $-w$ *and the necessary derivatives exist, then the quantities defined by*

$$E^{\alpha.\beta b.\gamma c} = \begin{bmatrix}\mathrm{A}, \mathrm{B}, \Gamma \\ M\end{bmatrix} T^{0bc[\mathrm{A}+\mathrm{B}+\Gamma-2M]}, \ \mathrm{A} = \alpha, \ \mathrm{B} = \beta, \ \Gamma = \gamma,$$

are the components of an extensor which is j-contravariant of order 1 *and excontravariant of order* 2.

Remark. To make the complete formal equivalence of this theorem with Theorem 2. 1 of [5] evident, we present for comparison the following

Proof. The quantity $(T^{0.bc}A_{M}B_{b}C_{c})^{(M)}$ is an invariant for arbitrary choice of the scalar of weight w, A_{M}, and the covariant vectors B and C.

---

[2] See reference [5] pages 334-336.

Dropping the original indices, performing the indicated differentiation, and finally replacing α with $M-\alpha$, etc. we get

$$(T\,A\,B\,C)^{(M)} = \sum \binom{M}{\alpha,\beta,\gamma} A^{(\alpha)} B^{(\beta)} C^{(\gamma)} T^{(M-\alpha-\beta-\gamma)}$$

$$= \sum \binom{M}{M-\alpha, M-\beta, M-\gamma} \binom{M}{\alpha}^{-1} \binom{M}{\beta}^{-1} \binom{M}{\gamma}^{-1} \binom{M}{\alpha} A_{M\cdot}{}^{(M-\alpha)} \binom{M}{\beta} B_{b\cdot}{}^{(M-\beta)}$$

$$\binom{M}{\gamma} C_{c\cdot}{}^{(M-\gamma)}\, T^{0.bc(\alpha+\beta+\gamma-2M)} = E^{a.\beta b.\gamma c} A_{a} M_{\beta b} C_{\gamma c},$$

and the theorem follows from the quotient laws. The summation on α, β, γ is from 0 to M throughout and the quantities A_{a}, $B_{\beta b}$, $C_{\gamma c}$ denote, respectively, the extensors $\binom{M}{\mathrm{A}} A_{M\cdot}{}^{(M-\mathrm{A})}$, $\binom{M}{\mathrm{B}} B_{b\cdot}{}^{(M-\mathrm{B})}$, $\binom{M}{\Gamma} C_{c\cdot}{}^{(M-\Gamma)}$.

THEOREM 4. 7. *If $T^{abc}{}_{M}$ is a third order contravariant tensor of weight w, then $E^{aa.\beta b.\gamma c}_{\delta}$, defined by the statements:*

$$E^{aa.\beta b.\gamma c}_{\delta} = \left\{ \begin{matrix} \mathrm{A}, \mathrm{B}, \Gamma \\ M, \Delta \end{matrix} \right\} T^{abc}{}_{M}{}^{(\mathrm{A}+\mathrm{B}+\Gamma-2M-\Delta)},$$

$$\left\{ \begin{matrix} \mathrm{A}, \mathrm{B}, \Gamma \\ M, \Delta \end{matrix} \right\} = \frac{\mathrm{A}!\mathrm{B}!\Gamma!}{M!M!\Delta!(\mathrm{A}+\mathrm{B}+\Gamma-2M-\Delta)!} \text{ or } 0$$

according as $\mathrm{A}+\mathrm{B}+\Gamma-2M-\Delta \geqq 0$ or < 0, is an extensor of excontravariant order three and j-covariant order one.

Remark. The multiplier of M in $(\mathrm{A}+\mathrm{B}+\Gamma-2M-\Delta)$ and the number of times $M!$ occurs in the denominator of the preceding fraction is *one* less than the number of doublet *superscripts* on E. To illustrate, if there are q top numbers in $\left\{ \begin{matrix} \mathrm{A}, \mathrm{B}, \cdots \\ M, \Gamma, \Delta \end{matrix} \right\}$, then the bottom number M should be regarded as occurring $(q-1)$ times, and thus the symbol $\{\ \}$ is equal to $(\mathrm{A}!\mathrm{B}!\cdots) \div (M!^{q-1}\Gamma!\Delta![\mathrm{A}+\mathrm{B}+\cdots-(q-1)M-\Gamma-\Delta]!)$.

THEOREM 4. 8. *If $T^{ab\cdots}_{M.d}$ is contravariant of order q, of weight w, and covariant of order one, then $E^{aa.\beta b\ldots}_{\gamma.\delta d}$ defined by*

$$E^{aa.\beta b\ldots}_{\gamma.\delta d} = \left\{ \begin{matrix} \mathrm{A}, \mathrm{B}, \cdots \\ M, \Gamma, \Delta \end{matrix} \right\} T^{ab\ldots[\mathrm{A}+\mathrm{B}+\ldots-(q-1)M-\Gamma-\Delta]}_{M.d}$$

is an extensor which is extravariant of order q, j-covariant of order one, and excontravariant of order one.

THEOREM 4. 9. *If $T_{M.bc}$ is a second order covariant tensor of weight w*

and if $E_{a.\beta b.\gamma c}$ *denotes* $\binom{M}{\mathrm{A},\mathrm{B},\Gamma}T^{\dots(M-\mathrm{A}-\mathrm{B}-\Gamma)}_{M.bc}$, *then* $E_{a.\beta b.\gamma c}$ *is an extensor of j-covariant order one and excovariant order two.*

Remark. These theorems are sufficiently typical to indicate the general situation. They may be proved directly or by following the method used in establishing Theorem 4. 6, or by differentiating the transformation equations of the given tensors by the product rule of Leibnitz. Perhaps it should be remarked in passing that the treatment of subscripts on T is much simpler than the treatment of superscripts. In the former case it is not necessary to change the associated Greek index, and to introduce the reciprocals of the binomial coefficients. To illustrate, the proof of Theorem 4. 9 may be read out of the following equation:

$$(T_{M.bc}A^0B^bC^c)^{(M)} = \sum\binom{M}{\alpha,\beta,\gamma}A^{0\,(\alpha)}B^{b\,(\beta)}C^{c(\gamma)}T^{\dots(M-\alpha-\beta-\gamma)}_{M.bc}.$$

Similarly, by expanding the indicated derivatives by means of the multi-product rule and proceeding as in the proofs of the foregoing theorems, we obtain the formulas:

$$(T^{0.bc}B_bC_c)^{(a)} = E^{a.\beta b.\gamma c}B_{\beta b}C_{\gamma c}, \tag{4. 4}$$

$$\binom{M}{\Delta}(T^{abc}{}_MA_aB_bC_c)^{(M-\Delta)} = E^{aa.\beta b.\gamma c}{}_\delta A_{aa}B_{\beta b}C_{\gamma c}, \tag{4. 5}$$

$$\binom{M}{\Gamma}(T^{ab\dots}_{M.d}\,A_aB_b\cdot\cdot\cdot D^d)^{(M-\Gamma)} = E^{aa.\beta b\dots}_{\gamma.\delta d}\,A_{aa}B_{\beta b}\cdot\cdot\cdot D^{d(\delta)}, \tag{4. 6}$$

$$\binom{M}{\mathrm{A}}(T_{M.bc}B^bC^c)^{(M-\mathrm{A})} = E_{a.\beta b.\gamma c}B^{b\,(\beta)}C^{c\,(\gamma)}. \tag{4. 7}$$

Finally, we note for future reference that if the extensors A_{aa}, $B^{b\,(\beta)}$ etc. appearing in the right members of these relationships are all expressible by extensor equations of the type $V^{a(a)} = V^bL^{aa}{}_b$ or $V_{aa} = V_bL^b{}_{aa}$, then we have at once:

$$(T^{0.bc}B_bC_c)^{(a)} = E^{a.\beta b.\gamma c}L^f{}_{\beta b}L^g{}_{\gamma c}B_fC_g; \tag{4. 8}$$

$$\binom{M}{\Delta}(T^{abc}{}_MA_aB_bC_c)^{(M-\Delta)} = E^{aa.\beta b.\gamma}{}_{c\delta}L^e{}_{aa}L^f{}_{\beta b}L^h{}_{\gamma q}A_eB_fC_h; \tag{4. 9}$$

$$\binom{M}{\Gamma}(T^{ab\dots}_{M.d}A_aB_b\cdot\cdot\cdot D^d)^{(M-\Gamma)} = E^{aa.\beta b\dots}_{\gamma.\delta d}\,L^e{}_{aa}L^f{}_{\beta b}\cdot\cdot\cdot L^{\delta d}{}_hA_eB_f\cdot\cdot\cdot D^h; \tag{4. 10}$$

$$\binom{M}{\mathrm{A}}(T_{M.bc}B^bC^c)^{(M-\mathrm{A})} = E_{a.\beta b.\gamma c}L^{\beta b}{}_dL^{\gamma c}{}_eB^dC^e. \tag{4. 11}$$

In our application of these formulas the quantities L will be taken to be the extended components of connection introduced in the major division immediately following.

Part B.

The foregoing developments have been presented against the background of a space devoid of the structure imposed by an affine connection, and are of course valid regardless of the nature of any connection that might be added subsequently. In the two sections which follow, numbers five and six, we shall apply the preceding theory to the fundamental problem of constructing tensors from weighted tensors by differentiation with respect to the curve parameter—or, in a word, the problem of intrinsic differentiation. Accordingly, we now assume that we have given in addition to our space a set of components of connection of class C^M and, in order to secure the possibility of introducing geodesic coordinates, we suppose further that these components of connection are symmetric three index symbols $L^a{}_{bc}$, $L^a{}_{bc} = L^a{}_{cb}$. From the quantities $L^a{}_{bc}$, we can of course derive the two index components of connection $L^a{}_b$, $L^a{}_b = L^a{}_{bc}x'^c$ and finally the extended components of connection $L^{\alpha a}{}_b$, $L^b{}_{\alpha a}$. These latter quantities may be constructed from the following formulas: $L^c{}_{\alpha a} = \binom{M}{\mathrm{A}} I^c{}_{\alpha a}$, $\mathrm{A} = \alpha$; $I^c{}_{\alpha-1\cdot a} = I^{c\prime}{}_{\alpha a} + I^b{}_{\alpha a}L^c{}_b$, $I^c{}_{Ma} = \delta^c{}_a$; $L^{\alpha+1.a}{}_c = L^{\alpha a\prime}{}_c - L^{\alpha a}{}_b L^b{}_c$, $L^{0a}{}_c = \delta^a{}_c$. The extended components of connection are extensors and are involved in iterated intrinsic differentiation. In addition, they enable us to express the derivatives of the components of a vector which is equipollent along a given curve in terms of the components of the vector. Specifically, if V^a and V_a are the components of vectors defined and equipollent along a curve C, then[3] $V^{a(\alpha)} = V^b L^{\alpha a}{}_b$ and $V_{\alpha a} = V_b L^b{}_{\alpha a}$, $V_{\alpha a} = \binom{M}{\mathrm{A}} V_a{}^{\cdot(M-\mathrm{A})}$, $\mathrm{A} = \alpha$, A not summed. Thus we note in passing that formulas (4.8)-(4.11) inclusive are valid if the vectors A, B, C are each equipollent along the curve in question. Perhaps it should be emphasized that equipollence throughout space is not required.

5. Jacobian connections. In addition to the foregoing methods for constructing Jacobian extensors from weighted scalars and tensors, it develops that it is possible to construct the new quantities from certain absolute extensors by merely contracting components through their Latin indices and then multiplying by the appropriate invariants, $1/N$ and w. To convince ourselves that such procedures may be fruitful, it will suffice to compare the transformation equations for Jacobian and absolute extensors with M equal to unity. These equations expanded on the Greek superscripts are:

---

[3] These equations were obtained by B. B. Townsend and H. V. Craig in a paper that has not yet been submitted for publication. They may be proved readily by induction.

$$J^{\rho}{}_{M} = J^{0}{}_{M}\underline{X}X^{\rho}{}_{0} + J^{1}{}_{M}\underline{X}X^{\rho}{}_{1}; \quad E^{\rho r}{}_{s} = E^{0a}{}_{b}X^{\rho r}{}_{0a}X^{b}{}_{s} + E^{1a}{}_{b}X^{\rho r}{}_{1a}X^{b}{}_{s}.$$

In either equation if $\rho = 0$, the second term drops out and the Jacobian equation becomes the transformation equation for an absolute invariant while the second equation reduces to the transformation equation of an absolute tensor of the right type to yield an invariant on contraction. When $\rho = 1$, the first equation becomes

$$(5.1) \qquad \bar{J}^{1}{}_{M} = J^{0}{}_{M}\underline{X}\bar{X}' + J^{1}{}_{M} = J^{0}{}_{M}wln'(\bar{x}/x) + J^{1}{}_{M},$$

while the second may be written

$$(5.2) \qquad E^{1r}{}_{s} = E^{0a}{}_{b}X^{r'}{}_{a}X^{b}{}_{s} + E^{1a}{}_{b}X^{b}{}_{s}X^{r}{}_{a}.$$

Now, according to a well known formula, $ln'(\bar{x}/x) = X^{r'}{}_{a}X^{a}{}_{r}$ ($X^{a}{}_{r}$ is the normalized cofactor of $X^{r}{}_{a}$). Consequently, if $E^{0d}{}_{b}$ is the Kronecker delta $\delta^{a}{}_{b}$ and we multiply equation (5.2) by w and contract, the result $wE^{1r}{}_{r} = wln'(\bar{x}/x) + wE^{1a}{}_{a}$ may be identified with (5.1), if we adopt the definitions: $\bar{J}^{1}{}_{M} = wE^{1r}{}_{r}$, $J^{1}{}_{M} = wE^{1a}{}_{a}$, $J^{0}{}_{M} = 1$ $(= 1/N\delta^{a}{}_{a} = 1/NE^{0a}{}_{a})$. To recapitulate, we assert

THEOREM 5.1. *If $E^{\alpha a}{}_{b}$ is an absolute extensor of the type indicated by its indices for $M = 1$ and if further $E^{0a}{}_{b} = \delta^{a}{}_{b}$, then the quantities $J^{\alpha}{}_{M}$ defined by $J^{0}{}_{M} = 1/NE^{0a}{}_{a}$, $J^{1}{}_{M} = wE^{1a}{}_{a}$, constitute the components of a Jacobian extensor.*

The corresponding proposition for extensors of the type $E^{b}{}_{\alpha a}$ is as follows:

THEOREM 5.2. *If $E^{b}{}_{\alpha a}$ is an extensor with $M = 1$, $E^{b}{}_{1a} = \delta^{b}{}_{a}$ then the quantities $J^{0}{}_{a}$ defined by*

$$J^{0}{}_{a} = \mathrm{I}(\mathrm{A})E^{a}{}_{\mathrm{A}a}, \qquad \mathrm{A} = \alpha, \qquad I(0) = W, \qquad I(1) = 1/N,$$

are the components of a Jacobian extensor.

The proof of this theorem is similar to that of number (5.1), accordingly we omit it.

Remark. The connection extensors $L^{\alpha a}{}_{b}$ and $L^{b}{}_{\alpha a}$, with $M = 1$, are examples of extensors of the types involved in the two preceding theorems.

These extensors it will be recalled are defined by the equations $L^{0a}{}_b = L^a{}_{1b} = \delta^a{}_b$, $L^{1a}{}_b = -L^a{}_b = -L^a{}_{bc}x'^c$, $L^a{}_{0b} = L^a{}_b$. Here the $L^a{}_{bc}$ are the Christoffel symbols or, more generally, the components of a symmetric connection. Symmetry is desirable since it allows the introduction of geodesic coordinates.

DEFINITION 5.1. *The Jacobian extensors* $J^a{}_M$ *and* $J^0{}_a$ *which are derived from the ordinary connection extensors in accordance with the formulas:*

$$J^a{}_M = I(\mathrm{A})L^{\mathrm{A}a}{}_a, \qquad J^0{}_a = I(\mathrm{A})L^a{}_{\mathrm{A}a}, \qquad \mathrm{A} = \alpha, \qquad M = 1,$$

with $I(\mathrm{A}) = 1/N$ *or* w *according as the index* A *on the operand* L *is a tensor index or a nontensor index, will be called the Jacobian connection extensors. The table of values for these symbols is* $J^0{}_1 = 1$, $-J^1{}_1 = J^0{}_0 = wL^a{}_a$.

Remark. It will be found that the Jacobian connection extensors play a rôle in the intrinsic differentiation of weighted tensors which is entirely analogous to that played by the ordinary connection extensors in the differentiation of absolute tensors.

As an illustration of this remark, let us consider the following proposition:

THEOREM 5.1. *If* S_M *and* S *are scalars of weights* w *and* $-w$, *respectively, and if* $S_a = \binom{M}{\mathrm{A}} S_{M\cdot}{}^{(M-\mathrm{A})}$, $S^a = S^{(a)}$, *then the contractions* $J^a{}_M S_a$ *and* $J^0{}_a S^a$, $M = 1$, *yield the intrinsic derivatives of* S_M *and* S.

Proof. $J^a{}_M S_a = J^0{}_M S_0 + J^1{}_M S_1, = S'_M - S_M w L^a{}_a = \delta S_M/\delta t$, and $J^0{}_a S^a = S' + SwL^a{}_a = \delta S/\delta t$.

Extensive differentiation and the extended Jacobian connection. A moment's consideration will show that the processes *upper* and *lower extensive differentiation*, which were defined in [4], have their counterparts in the present theory. Thus, if $E^0{}_\delta$ is a Jacobian extensor of the type indicated by its indices, S^0 is a scalar of weight $-w$, while S^δ denotes $S^{0(\delta)}$, then $E^0{}_\delta S^\delta$ and its intrinsic derivative $I(E^0{}_\delta S^\delta)$ are scalars of weight $-w$—(the operator I indicates intrinsic differentiation). Consequently, by the quotient law the multipliers of S^δ in $I(E^0{}_\delta S^\delta)$ constitute the components of a Jacobian extensor of range $M+1$. We denote this derived extensor by the symbol $DE\,|^0{}_\delta$ and call it the *upper extensive* derivative of $E^0{}_\delta$. The details of this procedure are as follows:

$$\begin{aligned} I(E^0{}_\delta S^\delta) &= E^0{}_\delta S^{\delta+1} + E^{0\prime}{}_\delta S^\delta + E^0{}_\delta wLS^\delta \\ &= \sum_{\delta=0}^{M} (E^0{}_{\delta-1} + E^{0\prime}{}_\delta + E^0{}_\delta wL)S^\delta + E^0{}_M S^{M+1}, \qquad L = L^a{}_a. \end{aligned}$$

Thus we are led to formulate

DEFINITION 5.2. *If* $E^0{}_\delta$ *is a Jacobian extensor, then the quantities* $DE|^0{}_\delta$ *given by the equations*

$$DE|^0{}_0 = E^{0\prime}{}_0 + E^0{}_0 wL;\quad DE|^0{}_\delta = E^0{}_{\delta-1} + E^{0\prime}{}_\delta + E^0{}_\delta wL,\ DE|^0{}_{M+1} = E^0{}_M,$$

will be called the upper extensive derivative of $E^0{}_\delta$.

Remark. The process *upper extensive differentiation* extends the range of the variable index. Except for the additional term, it is similar to intrinsic differentiation of a contravariant vector.

The *lower extensive derivative* may be obtained by a similar procedure. The starting point is the intrinsic derivative of the weighted scalar $E^a{}_M S_a$. Here $S_a = \binom{M}{A} S_M{}^{\cdot(M-A)}$ with S_M a scalar of weight w. The intrinsic derivative of $E^a{}_M S_a$ will involve linearly the derivatives of S_M from order zero to order $M+1$. Application of the quotient law will then yield a new extensor which involves the total derivative of E and the connection quantity L. In order to carry out this procedure it will be necessary first to evaluate S_a and S'_a in terms of $S^*{}_a$, the corresponding quantities for range $M+1$, with $S^*{}_{M+1} = S_M$. These formulas

$$(5.3)\qquad (M+1)S'_a = (M-A+1)S^*{}_a,\ (M+1)S_a = (A+1)S^*{}_{A+1},\quad A = \alpha,$$

may be verified readily. Applying them to the intrinsic derivative, which is given by the equality,

$$I(E^a{}_M S_a) = E^a{}_M S'_a + E^{a\prime}{}_M S_a - E^a{}_M S_a wL,$$

we get the relationship

$$\begin{aligned}(M+1)I(E^a{}_M S_a) &= \sum_{\alpha=0}^{M}(M-\alpha+1)E^a{}_M S^*{}_a \\ &\quad + \sum_{\alpha=1}^{M+1}\alpha(E^{a-1\prime}{}_M - E^{a-1}{}_M wL)S^*{}_a \\ &= \sum_{\alpha=0}^{M+1}[(M-\alpha+1)E^a{}_M + \alpha(E^{a-1\prime}{}_M - E^{a-1}{}_M wL)]S^*{}_a.\end{aligned}$$

The conclusion is that the quantities in the bracket constitute the components of an extensor of range zero to $M+1$.

DEFINITION 5.3. *If* $E^a{}_M$ *is a Jacobian extensor of the type indicated by the free indices then the quantities* $D_1E|^a{}_{M+1}$ *defined by* $D_1E|^0{}_{M+1} = E^0{}_M$; $(M+1)D_1E|^a{}_{M+1} = (M-\alpha+1)E^a{}_M + \alpha(E^{a-1\prime}{}_M - E^{a-1}{}_M wL)$ *will be called the lower extensive derivative of the original extensor* $E^a{}_M$.

Remark. Again the range of the derived extensor is *one greater* than that of the original extensor.

Since any absolute invariant is a Jacobian extensor of the type $E^0{}_M$, it follows that we may apply *upper* and lower *extensive differentiation* to any such quantity. As an illustration, we complete the extensive derivatives of unity. Thus, if we take the $M = 0$ and $E^0{}_M = 1$ and apply Definition 5.2, we get $DE|^0{}_1 = 1$; $DE|^0{}_0 = wL$, the components of the Jacobian connection extensor $J^0{}_\alpha$. Similarly if we apply *lower extensive differentiation* (Definition 5.3), we obtain the results $D_1E|^0{}_1 = 1$; $D_1E|^1{}_1 = -wL$ or $D_1E|^\alpha{}_1 = J^\alpha{}_1$. As a recapitulation, we state

THEOREM 5.2. *The upper and lower extensive derivatives of unity are equal respectively to $J^0{}_\alpha$ and $J^\alpha{}_1$, the components of the Jacobian connection extensors.*

From the development of extensive differentiation, it is obvious that if $E^0{}_\alpha S^\alpha$ gives the M-th intrinsic derivative of S^0, then $DE|^0{}_\alpha S^\alpha$, with α summed from zero to $M+1$, will give the intrinsic derivative of order $M+1$ of S^0. Likewise, $D_1E|^\alpha{}_{M+1}S^*{}_\alpha$ with α summed from 0 to $M+1$ is the intrinsic derivative of $S^*{}_{M+1}$ ($S^*{}_{M+1} = S_M$) of order $M+1$, if $E^\alpha{}_M S_\alpha$ is the M-th order intrinsic derivative of S_M. Consequently, since the first order extensive derivatives of unity constitute the components of the Jacobian connection extensor, we may generate quantities E, having the properties just postulated, by repeated extensive differentiation. With this as a background, we introduce

DEFINITION 5.4. *The Jacobian extensor generated by applying repeatedly upper and lower extensive differentiation to unity will be denoted by the symbols $J^0{}_\alpha$ and $J^\alpha{}_M$, respectively, and will be referred to as the extended Jacobian connections;*

and assert

THEOREM 5.3. *If S is a scalar of weight $-w$ and S^α denotes $S^{(\alpha)}$, then the M-th order intrinsic derivative of S is given by the relationship $I^M S = J^0{}_\alpha S^\alpha$;*

and

THEOREM 5.4.[4] *If S_M is a scalar of weight w and S_α denotes*

---

[4] Mr. J. M. Hurt has treated in his Master's thesis the problem of expressing certain intrinsic derivatives of weighted tensors as contractions of the appropriate extensors with the components of the absolute extended connection. In particular he investigated the case of the first order derivatives of higher order weighted tensors and the first and second derivatives of weighted vectors.

$\binom{M}{A} S_{M}\cdot^{(M-A)}$, *then the contraction* $J^{a}{}_{M}S_{a}$ *yields the M-th order intrinsic devrivative of* S_M, $I^{M}S_{M}$.

Because of the formal similarity of Jacobian extensive differentiation and absolute extensive differentiation, we may take over the rules given in [3] for the computation of extended connections. Specifically, to compute $J^{0}{}_{a}$, we proceed as follows: We start with $J^{0}{}_{M}=1$; then we differentiate 1 and add in the product of 1 with wL; next we take the result just obtained and differentiate it and add in its product with wL. This process is continued until we have constructed $M+1$ terms. The final step is to multiply in the binomial coefficients from row number $M+1$ of Pascal's triangle.

The procedure to be followed in the construction of $J^{a}{}_{M}$ is somewhat simpler. We start with $J^{0}{}_{M}=1$ as before, but instead of adding the products with wL we subtract these products and do not introduce the binomial coefficients. Thus to obtain any component of $J^{a}{}_{M}$ after the given one, we differentiate the preceding component and subtract the product of this preceding component with wL.

As an illustration of these rules, we compute $J^{0}{}_{a}$ and $J^{a}{}_{3}$ for the case $M=3$. The calculation is as follows:

$$J^{0}{}_{3}=1,\ J^{0}{}_{2}=3(1'+1wL)=3wL,\ J^{0}{}_{1}=3(wL'+w^{2}L^{2}),$$
$$J^{0}{}_{0}=(wL'+w^{2}L^{2})'+(wL'+w^{2}L^{2})wL;$$

$$J^{0}{}_{3}=1,\ J^{1}{}_{3}=1'-wL=-wL,\ J^{2}{}_{3}=-wL'+w^{2}L^{2},$$
$$J^{3}{}_{3}=(-wL'+w^{2}L^{2})'-(-wL'+w^{2}L^{2})wL.$$

6. The structure of the intrinsic derivatives of weighted tensors. In a preceding paper [5], the M-th order intrinsic derivatives of certain typical higher order tensors are expressed as contractions of the extended components of connection with certain derived extensors. Thus it appears that these rather complicated derivatives may be "factored" into quantities which themselves have desirable invariantive properties. We shall now show that the M-th order intrinsic derivatives of higher order *weighted* tensors admit of a similar "factorization."

The method of attack is to contract the weighted tensors with arbitrary equipollent vector fields so as to yield a weighted scalar, which we denote by S_M or S^0 according as the weight is $+w$ or $-w$, $w>0$. We then form the M-th order intrinsic derivative of S first employing Theorem 5. 4 or Theorem 5. 3 and then by employing the analogue for intrinsic derivatives of the formula for the derivative of a multiproduct. This formula may be verified readily by introducing geodesic coordinates and considering the key case $M = 1$. The extension to higher values of M follows the pattern of the parent theorem.

The specific theorems to be studied are:

THEOREM 6. 1. *If $T^{0.ab}$ is a tensor of weight $-w$ and of contravariant order two and if $E^{a.\beta b.\gamma c}$ is the associated extensor* (given in Theorem 4. 6), *then $J^0{}_a E^{a.\beta b.\gamma c} L^f{}_{\beta b.} L^g{}_{\gamma c}$ is the M-th order intrinsic derivative of $T^{0.ab}$:*

Proof. Let S^0 denote $T^{0.bc} B_b C_c$, with B and C arbitrary, equipollent, absolute vectors. By way of Theorem 5. 3 and Equation 4. 8, we may write

$$I^M S^0 = J^0{}_a S^{0.(a)} = J^0{}_a E^{a.\beta b.\gamma c} L^f{}_{\beta b} L^g{}_{\gamma c} B_f C_g.$$

Similarly, from the rule for expressing derivatives of a multiproduct in terms of the derivatives of the factors, and the fact the intrinsic derivatives of B and C of order greater than zero vanish, we have

$$I^M S^0 = \sum \binom{M}{\beta, \gamma} (I^\beta B_b)(I^\gamma C_c) \; I^{M-\beta-\gamma} T^{0.bc} = B_f C_g I^M T^{0.fg}.$$

Consequently,

$$B_f C_g I^M T^{0.fg} = B_f C_g J^0{}_a E^{a.\beta b.\gamma c} L^f{}_{\beta b} L^g{}_{\gamma c},$$

and the theorem follows by way of the arbitrariness of B and C.

THEOREM 6. 2. *If $T^{efg}{}_M$ is a tensor of weight w and of contravariant order three, and if $E^{\alpha a.\beta b.\gamma c}{}_\delta$ is the corresponding derived extensor, then $J^\delta{}_M E^{\alpha a.\beta b.\gamma c}{}_\delta L^e{}_{\alpha a} L^f{}_{\beta b} L^g{}_{\gamma c}$ is the M-th order intrinsic derivative of $T^{efg}{}_M$.*

Proof. Following the pattern of the previous proof, we employ Theorem 5. 4 and Equation 4. 9 and conclude that

$$I^M S_M = J^\delta{}_M S_\delta = J^\delta{}_M \binom{M}{\delta} S_{M}{}^{\cdot(M-\delta)} = J^\delta{}_M E^{\alpha a.\beta b.\gamma c}{}_\delta L^e{}_{\alpha a} L^f{}_{\beta b} L^g{}_{\gamma c} A_e B_f C_g.$$

Next we apply the multiproduct rule to S_M and simplify the result by taking advantage of the equipollence of the vectors A, B, C. Thus we have

$$I^M S_M = \sum \binom{M}{\alpha, \beta, \gamma} (I^\alpha A_a)(I^\beta B_b)(I^\gamma C_c) I^{M-\alpha-\beta-\gamma} T^{abc}{}_M = A_a B_b C_c I^M T^{abc}{}_M.$$

The theorem then follows by comparison of results.

THEOREM 6. 3. *If $T^{ab\cdots}{}_{M.d}$ is contravariant of order q, of weight w, and covariant of order one, and $E^{\alpha a.\beta b\cdots}{}_{\gamma.\delta d}$ is the extensor presented in* Theorem

4.8, *then the contraction* $J^{\gamma}{}_{M}E^{\alpha a.\beta b\ldots}{}_{\gamma.\delta d}L^{e}{}_{\alpha a}L^{f}{}_{\beta b}\cdots L^{\delta d}{}_{h}$ *is the M-th order intrinsic derivative of* $T^{ef\ldots}{}_{M.h}$.

Proof. Since the salient features of the proof are the same as in the two preceding theorems, we present the argument in skeleton form only. The foundations are supplied by Theorem 5.4 and Formula 4.10. Briefly, we have, on the one hand,

$$I^{M}S_{M}=J^{\gamma}{}_{M}S_{\gamma}=J^{\gamma}{}_{M}\binom{M}{\gamma}S_{M^{\cdot}}{}^{(M-\gamma)}=J^{\gamma}{}_{M}E^{\alpha a.\beta b\ldots}{}_{\gamma.\delta d}L^{e}{}_{\alpha a}L^{f}{}_{\beta b}L^{\delta d}{}_{h}A_{e}B_{f}D^{h};$$

while on the other

$$\begin{aligned}I^{M}S_{M}=\sum\binom{M}{\alpha\beta\delta}(I^{\alpha}A_{a})(I^{\beta}B_{b})\cdots(I^{\delta}D^{d})I^{M-\alpha-\beta\ldots-\delta}T^{ab\ldots}{}_{M.d}\\ =A_{e}B_{f}\cdots D^{h}I^{M}T^{ef\ldots}{}_{M.h}.\end{aligned}$$

THEOREM 6.4. *If* T_{Mbc} *is a second order covariant tensor of weight* w *and* $E_{\alpha.\beta b.\gamma c}$ *is the extensor of* Theorem 4.9, *then*

$$J^{\alpha}{}_{M}E_{\alpha.\beta b.\gamma c}L^{\beta b}{}_{d}L^{\gamma c}{}_{e}=I^{M}T_{M.de}.$$

Proof. From Formula 4.11 and procedures similar to those employed in establishing the foregoing theorems, we get the equalities:

$$I^{M}S_{M}=J^{\alpha}{}_{M}S_{\alpha}=J^{\alpha}{}_{M}\binom{M}{\alpha}S_{M^{\cdot}}{}^{(M-\alpha)}=J^{\alpha}{}_{M}E_{\alpha.\beta b.\gamma c}L^{\beta b}{}_{d}L^{\gamma c}{}_{e}B^{d}C^{e},$$

and

$$I^{M}S_{M}=\sum\binom{M}{\beta\gamma}(I^{\beta}B^{b})(I^{\gamma}C^{c})I^{(M-\beta-\gamma)}T_{M.bc}=B^{d}C^{e}I^{M}T_{M.de}$$

and the theorem is established.

UNIVERSITY OF TEXAS.

BIBLIOGRAPHY.

1. A. Kawaguchi, "Die Differentialgeometrie höherer Ordnung I. Erweiterte Koordinantentransformationen und Extensoren," *Journal of the Faculty of Science, Hokkaido Imperial University* (I), vol. 9 (1940).
2. H. V. Craig, "On tensors relative to the extended point transformation," *American Journal of Mathematics*, vol. 59 (1937), pp. 764-774.
3. ———, *Vector and Tensor Analysis*, New York (1943).
4. ———, "On the structure of certain tensors," *Mathematics Magazine*, vol. I (1947), pp. 21-30.
5. ———, "On the structure of intrinsic derivatives," *Bulletin of the American Mathematical Society*, vol. 53 (1947), pp. 332-342.

ALGEBRAIC SYSTEMS OF POSITIVE CYCLES IN AN ALGEBRAIC VARIETY.*

By WEI-LIANG CHOW.

1. Introduction.

In a paper[1] some years ago van der Waerden and the author introduced the idea of the associated form of a positive cycle. For a positive cycle of dimension d and degree n in the projective space S_m of m dimensions the associated form is a form of $d+1$ sets of $m+1$ indeterminates each, which has the degree n in each set of the indeterminates. If we arrange all the monomials of this type in an arbitrary but fixed order $\omega_0, \omega_1, \cdots, \omega_t$, then each form of this type can be expressed in a unique way as a linear combination $\sum_{i=0}^{t} c_i\omega_i$ of the monomials. By assigning to each such form the point in the projective space S_t with the (homogeneous) coordinates $(c_0, c_1, \cdots, c_t)$, we obtain a one-to-one correspondence between all the forms of this type and all the points of S_t. Applying this in particular to the associated form, we can then assign to each positive cycle of dimension d and degree n in S_m a point in S_t; we shall call the coordinates of this point, which are the coefficients of the associated form, the natural coordinates of the positive cycle. In this way the totality of all the positive cycles of dimension d and degree n in S_m (or in any variety contained in S_m) is represented by a set of points in S_t. The main theorem about this representation of positive cycles in S_m by points in S_t, proved in the above mentioned paper,[2] asserts that this set of points constitutes a bunch of varieties in S_t. An important consequence of this main theorem is that it enables us to give a precise definition of an (irreducible) algebraic system of positive cycles in S_m: If V is a variety contained in the above mentioned bunch of varieties in S_t, then the set of all the positive cycles in S_m which correspond to the points in V is called an algebraic system of positive cycles. The variety V is then called the associated variety of the algebraic system. The associated variety gives a representation of the algebraic system of positive cycles in the sense that there is a one-to-one correspondence between the

---

* Received January 20, 1949.

[1] Chow and van der Waerden [5]. Following Weil [8], we shall use the expression "positive cycle" instead of "reducible variety."

[2] Chow and van der Waerden [5], p. 698 and p. 700.

points of V and the positive cycles of the algebraic system, and that this correspondence is of an algebraic nature. This latter statement means that there is an algebraic correspondence T between the variety V and the carrier variety U of the algebraic system[3] such that to each point of V there correspond under T exactly the points of the corresponding positive cycle. This algebraic correspondence T is called the associated correspondence of the algebraic system of positive cycles. The significance of such a representation of an algebraic system of positive cycles as an algebraic variety is easily seen, for it enables us to apply the concepts and results in the theory of algebraic varieties to the study of the algebraic systems. Thus, for example, the set of all linear subspaces (i. e. positive cycles of degree 1) of a given dimension in a projective space is represented by means of their natural coordinates, which in this special case are the well-known Pluecker coordinates, as a certain variety called the Grassmann variety, and various algebraic systems of such linear subspaces are then represented as subvarieties of the Grassmann variety. It is well-known that the study of these algebraic systems of linear subspaces is greatly facilitated by their representation as such subvarieties.

However, except in very special cases such as the Grassmann variety and its subvarieties just mentioned, the usefulness of the associated variety of an algebraic system has been hitherto rather limited. The main reason for this lies in the fact that we know very little about the properties of the associated variety besides its algebraic nature and the fact that it represents the system of positive cycles in a one-to-one manner. In fact, if we consider an associated variety in general without any restriction, it cannot have by itself any special properties; for it is easily seen that any variety whatsoever can be the associated variety of a suitably chosen algebraic system of positive cycles of any given dimension. Therefore we must look for whatever special properties that might exist in the relation between the carrier variety U and the associated variety V under the associated correspondence T; in other words, we must study the special properties of the associated correspondence T and the manner in which it connects the properties of U with those of V. In particular, it is important for the purpose of application to know under what conditions about the variety U and the correspondence T can we be sure that the associated variety V is non-singular or at least simple at a given point. For, most of the deeper concepts and results in the geometry of an algebraic variety can be applied without serious restrictions only to a non-singular variety or only to a simple point or subvariety of a variety. For example, the entire intersection theory has been up to now developed only for the case when the ambient variety is simple at the intersection point or variety, and there are strong indications that such an intersection theory cannot be extended to a substantially more general class of varieties without a radical change of its character. Thus the usefulness of the Grassmann variety in the study of the geometry of systems of linear subspaces is due by no mean measure to its being non-singular. A particular case in point is the problem of the Jacobian variety of an algebraic curve, a problem which we have studied in some detail in a recent paper.[4] We have shown there that each class of equivalent positive divisors of degree $n > 2g - 2$ on a non-singular curve C (where g is the genus of C) can be represented as a subvariety of dimension $n - g$ on the n-fold symmetric product C^n of the curve C, and that the set of all such subvarieties forms an algebraic system. The associated variety of this algebraic system would then be a Jacobian variety of the curve C, if we could prove that it is non-singular; and this requirement of being non-singular is essential, for without this property a Jacobian variety would not be of much use in the study of the geometry of the divisor classes. In the quoted paper we have solved the problem by proving that any derived normal model of the associated variety is a Jacobian variety; however, as we have indicated there, the entire results would be much more satisfactory if we could prove that the associated variety itself is already a Jacobian variety. That this is so will follow from the results of the present paper; in fact, the attempt to prove this particular result about the Jacobian variety is the original problem from which we started out and which leads us finally to the general problem considered here.

---

[3] The carrier variety of an algebraic system is defined as the smallest variety which contains all the cycles of the system.

Before we go over to the exact formulation of the problem and results, we shall give a short summary of the main definitions so that there will be no ambiguity about the terminology we are going to use in this paper. This is the more necessary in view of the fact that there has been so much divergence in terminology in the recent works in the field of algebraic geometry. Although this is a rather unfortunate state of affairs, yet it is in a sense not entirely avoidable; for the different terminologies of different authors are actually more suitable for their particular purposes, and it is difficult to adopt a uniform terminology without its being too cumbersome to be convenient. In fact, we ourselves are forced to adopt in the present paper a somewhat different terminology from that of our recent paper on the Jacobian variety; for, in our present problem the consideration of "relative" varieties over a given ground field is natural and in a sense essential, while in the other paper we have only to deal with the absolute varieties.

---

[4] Chow [4].

We consider a field $\mathfrak{K}$ called the "universal domain" which is an algebraically closed field of infinite degree of transcendency over the prime field contained in it. The set of all subfields K of the universal domain $\mathfrak{K}$ such that $\mathfrak{K}$ has an infinite degree of transcendency over K, has evidently the property that the intersection of any number of fields of the set is also a field of the set, and that any subfield of $\mathfrak{K}$ which is a finite extension of any field of the set is also a field of the set. In this paper, whenever we speak of "a field," we shall always mean a field of this set; on the few occasions when we have to consider a field not in this set, we shall expressly say so and refer to such a field as an "abstract field." Any ordered set of $m+1$ elements $(x)=(x_0, x_1, \cdots, x_m)$ in the universal domain $\mathfrak{K}$, not all zero, is called a point of the projective space S_m of dimension m, and two points are regarded as identical if their coordinates are proportional. Given any point (x) in S_m and a field K, the field $K((x))$ is the extension of K generated by the adjunction of the ratios of the $m+1$ coordinates $x_0, x_1, \cdots, x_m$ of the point (x). We can regard this field $K((x))$ as the field generated by the adjunction of the point (x); it serves the same purpose as the ordinary field extension by adjunction of coordinates in the case of a point in the affine space. In similar way we can define the extension $K((x),(y),\cdots)$ generated by the adjunction of several points $(x),(y),\cdots$ in the same or different projective spaces. The degree of transcendency of the field $K((x))$ over K is called the dimension of the point (x) over K; the point (x) is called algebraic over K if its dimension over K is zero, and it is called rational over K if $K((x))=K$. Beside the elements of the universal domain $\mathfrak{K}$, we shall consider also elements which are not in $\mathfrak{K}$, mainly as symbols by means of which we can express polynomials and functions. Such an element is called an indeterminate, and an ordered set of $m+1$ indeterminates $(X)=(X_0, X_1, \cdots, X_m)$ which are independent variables over $\mathfrak{K}$ is called a set of indeterminates in S_m. We shall always denote the indeterminates by such capital Latin letters as X, Y, Z.

Let K be a field and let (X) be a set of indeterminates in S_m. Any homogeneous prime ideal $\mathfrak{J}$ in the ring $K[X]=K[X_0, X_1, \cdots, X_m]$ is said to define a variety U/K in S_m over the field K. A point (x) is said to be in the variety U/K if all the polynomials of the ideal $\mathfrak{J}$ vanish for $(X)=(x)$. A variety W/K' over any other field K' is said to be contained in the variety U/K if every point of W/K' is in U/K; if we have furthermore $K=K'$, then the variety W/K is said to be a subvariety of U/K. Two varieties are identical if they are subvarieties of each other. Any point (x) in S_m determines over any field K a homogeneous prime ideal consisting of all the polynomials of $K[X]$ which vanish for $(X) = (x)$, and hence it determines a variety over K. A point (x') is called a specialization of the point (x) over K if it is contained in the variety determined by the point (x) over K; and if the point (x) is at the same time also a specialization of (x') over K, then the two points are called generic specializations of each other over K. Any variety U/K contains a point (x) such that the variety determined by (x) over K is identical with U/K; such a point (x) is called a generic point of U/K. Any two generic points (x) and (x') of a variety U/K are equivalent over K in the sense that the fields $K((x))$ and $K((x'))$ are K-isomorphic under the correspondence $(x) \leftrightarrow (x')$, i. e. the ratios of the corresponding coordinates are to correspond to each other. In particular, all generic points of a variety U/K have the same dimension over K; this dimension is called the dimension of the variety U/K. A finite set of varieties is called a bunch of varieties; and any point or variety contained in any one of the varieties of the bunch is said to be contained in the bunch.

Let K be a field containing K. If U/K is a variety such that its defining ideal $\mathfrak{J}$ in $K[X]$ generates also a prime ideal $\breve{\mathfrak{J}}$ in $\breve{K}[X]$, then the variety defined by $\breve{\mathfrak{J}}$ over K is called the extension U/K of U/K over K. The extension of a variety is in a sense essentially the same as the original variety, for they not only are contained in each other but also can be defined by the same set of equations.[5] This is the reason why we denote the extension by the same symbol U as the original variety. If a subvariety W/K of U/K has also an extension W/K over $\breve{K}$, then $W/\breve{K}$ is also a subvariety of $U/\breve{K}$. In order that a variety U/K has an extension $U/\breve{K}$ over $\breve{K}$, it is necessary and sufficient that the fields $\breve{K}$ and $K((x))$ are linearly disjoint[6] over K, where (x) is a generic point of U/K, free with respect to $\breve{K}$ over K.

Two varieties in S_m are called equivalent if they are both extensions of a third variety. It is easily seen that equivalent varieties are contained in each other and can be defined by the same set of equations; and that two varieties over the same field can only be equivalent if they are identical. We shall consider equivalent varieties as essentially the same and shall denote all of them by the same symbol over different fields such as $U/K, U/K', \cdots$. A variety over K which has an extension over any field containing K is called absolutely irreducible. If a variety is absolutely irreducible, then any variety equivalent to it is also absolutely irreducible. A complete set of mutually equivalent absolutely irreducible varieties is called an absolute variety, and any variety in the set is to be regarded as a representative of absolute variety over the particular field. If an absolute variety U has a representative U/K over a field K, then it is said to be defined over K. It is known [7] that among all the fields over which a given absolute variety U is defined, there is a smallest one which is contained in all of them; this smallest field is called the defining field of U. As an absolute variety is uniquely determined by any one of its representatives, there is little danger of confusion in identifying it sometimes with one of the latter. The absolute varieties are of great importance for us, for it is by means of them that we are going to define the concept of a cycle in S_m. Before doing this, we shall introduce the associated form of a variety.

---

[5] Following Weil [8], we shall say that a variety is defined by a set of equations $f_i(X) = 0$, if the polynomials $f_i(X)$ generate the corresponding prime ideal in $K[X]$.

[6] See Weil [8], p. 15, Theorem 3. We observe here that the prime ideal determined by the point (x) in $\breve{K}[X]$ has a basis in $K[X]$ if and only if the prime ideal determined by (x) in $K[X]$ generates a prime ideal in $\breve{K}[X]$.

Let U/K be a variety of dimension d in S_m. As we have shown elsewhere,[8] there is associated with U/K a K-irreducible form $F(Z, Z^{(1)}, \cdots, Z^{(d)})$ of the same degree n in each of the $d+1$ sets of indeterminates (Z), $(Z^{(1)})$, $\cdots$, $(Z^{(d)})$ in S_m with coefficients in K. This form is called the associated form of the variety U/K, and the number n is called the degree of the variety. Conversely, given any K-irreducible form F of this type satisfying a certain set of algebraic conditions, there is a uniquely determined variety U/K of which F is the associated form. It is easily seen that equivalent varieties have the same associated form; the converse of this statement is also true in case the varieties are separably generated or (which is the same thing) the associated form has no multiple factors. A variety is absolutely irreducible if and only if its associated form is absolutely irreducible. It follows then that an absolute variety is uniquely determined by its associated form; this is the reason why the associated forms are particularly suitable for the study of absolute varieties and cycles.

A cycle of dimension d or d-cycle in S_m is an element of the free Abelian group, the generators of which are the set of all absolute varieties of dimension d in S_m. Thus a d-cycle is a finite set of absolute varieties U_j, called its components, to each of which is assigned an integer n_j which is called the multiplicity of U_j in the d-cycle. If the degrees of the varieties U_j are s_j respectively, then the number $n = \sum_j s_j n_j$ is called the degree of the d-cycle. A cycle is said to be contained in a variety if each of its components is contained in the variety. A cycle is called positive if all the multiplicities n_j

---

[7] See Weil [8], pp. 70-1, Theorem 1, Corollary 3.

[8] Chow and van der Waerden [5], pp. 693-694.

are positive. If F_j are the associated forms of the varieties U_j respectively, then the form $F = \prod_j F_j^{n_j}$ is called the associated form of the positive cycle. A positive cycle is said to be rational over a field K if its associated form is rational over K, i. e. the ratios of its coefficients are rational over K; and if furthermore the associated form is irreducible over K, then the positive cycle is said to be prime rational[9] over K. Since such a K-irreducible form determines a variety over K and vice versa, it folows that there is a one-to-one correspondence between the varieties over K and the prime rational cycles over K. It is clear that every rational cycle over K can be factorized in a unique way into a sum of prime rational cycles over K corresponding to the factorization of its associated form into a product of irreducible forms over K. Therefore every rational cycle over K can also be regarded as a sum of varieties over K, each with a certain multiplicity; we shall call these varieties the K-component of the cycle.

Let (x) be a generic point of a variety U/K, then the field $K((x))$, which is uniquely determined up to an isomorphism by the variety, is called the field of U/K. Let (ξ) be any point of U/K, then the ring of all the elements of $K((x))$ of the form $f(x)/g(x)$, where $f(x)$ and $g(x)$ are two homogeneous polynomials of (x) of the same degree and $g(\xi) \neq 0$, is called the quotient ring $\mathfrak{U} = Q(U/K, (\xi))$ of U/K at the point (ξ). It is clear that if (ξ') is a generic specialization of (ξ) over K, then $Q(U/K, (\xi)) = Q(U/K, (\xi'))$, so that the ring $\mathfrak{U}$ depends only on the subvariety W/K in U/K determined by the point (ξ) or (ξ') over K, not on the particular choice of the generic point (ξ) or (ξ') of W/K. Therefore we shall also call the ring $\mathfrak{U}$ the quotient ring of U/K at the subvariety W/K, and shall sometimes write $\mathfrak{U} = Q(U/K, W)$ to indicate its dependence on W/K. The quotient ring $\mathfrak{U}$ is a local ring in the sense of Krull,[10] i. e. a Noetherian ring in which the set of all non-units is an ideal, and its maximal prime ideal $\mathfrak{u}$

---

[9] This is not the "correct" way to define a rational cycle, but is more convenient for our present purpose. The "correct" definition is as follows. A cycle is said to be rational over a field K if it satisfies the following conditions: (1) Each component of the cycle is algebraic over K; (2) if an (absolute) variety occurs in the cycle, then all the conjugate varieties over K also occur in it with the same multiplicity; (3) the multiplicity of each component is a multiple of its order of inseparability. See Weil [8], Chapter VII, § 6. Our definition differs from this "correct" definition in that the multiplicity of a component will in general be a multiple of a smaller power of the characteristic of K than the order of inseparability of this component over K.

[10] Krull [7]. Our Appendix, On the Extensions of Local Domains, contains a detailed treatment of many of the concepts and results concerning local rings which are used in the text.

consists of all elements $f(x)/g(x)$ such that $f(\xi) = 0$. It is also known[11] that the completion $\mathfrak{U}^*$ of the local ring $\mathfrak{U}$ has no nilpotent elements; in other words, the zero ideal in $\mathfrak{U}^*$ is an intersection of a finite number of prime ideals. Let V/K be another variety, and let $\mathfrak{B}$ be the quotient ring $Q(V/K, (\eta))$ of V/K at one of its points (η); if W/K is the subvariety in V/K determined by the point (η) over K, then $\mathfrak{B}$ is also the quotient ring $Q(V/K, W')$ of V/K at the subvariety W'/K. We shall say that the varieties U/K and V/K are analytically equivalent[12] at the points (ξ) and (η) respectively, or at the subvarieties W/K and W'/K respectively, if the completions $\mathfrak{U}^*$ and $\mathfrak{B}^*$ of the two local rings $\mathfrak{U}$ and $\mathfrak{B}$ are K-isomorphic. Local properties of a variety at a point or a subvariety which are invariant under analytical equivalences are called locally analytic properties. It is known[13] that the property of a variety being simple at a point or subvariety in the absolute sense is a locally analytic property.

For an absolute variety U the concept of a quotient ring can only be defined with respect to a given field of definition. However, in order that this concept has a significance independent of the choice of the particular field of definition, it is advisable to restrict ourselves to the case where the subvariety W is also an absolute variety and the field K is a field of definition for both U and W. Then the ring $Q(U/K, W)$ is called the quotient ring of the absolute variety U at the absolute subvariety W over the field K. This includes in particular the case when W has the dimension 0 and hence consists of a point which is rational over K. The advantage for imposing these restrictions regarding W and K lies in the fact that if K' is any field containing K, then the quotient ring $Q(U/K', W)$ is the extension[14] of the quotient ring $Q(U/K, W)$ over K', so that the ring $Q(U/K', W)$ is uniquely determined by the ring $Q(U/K, W)$ and the field K'. Furthermore, the completion of $Q(U/K', W)$ is the completion of the extension of the completion of $Q(U/K, W)$ over K', so that the completion of $Q(U/K', W)$ is also uniquely determined by the completion of $Q(U/K, W)$ and the field K'. Let V be another absolute variety and W' be an absolute subvariety in V; let K_0 be the smallest field over which all the varieties U, V, W, W' are defined. Then the absolute varieties U and V are said to be analytically equivalent at the subvarieties W and W' respectively, if the completions of the two local rings $Q(U/K_0, W)$ and $Q(V/K_0, W')$ are K_0-isomorphic. It follows then from what we have just said above that if K is any field containing K_0, then the completions of the two local rings $Q(U/K, W)$ and $Q(V/K, W')$ are also K-isomorphic; or, in other words, the (relative) varieties U/K and V/K are analytically equivalent at the subvarieties W/K and W'/K respectively.

---

[11] See Chevalley [3], p. 11, Theorem 1.

[12] In this definition of analytical equivalence we are following Zariski [10], p. 49, Definition 4.

[13] See Zariski [10], pp. 49-51, Theorem 15, where this theorem is proved for the most general case.

[14] See Appendix.

Let U/K and V/K be two varieties in the space S_m and S_t respectively, and let (x) be a generic point of U/K. If V/K has a generic point (y) which is rational over $K((x))$, then the correspondence $(x) \rightarrow (y)$ defines a rational transformation T of U/K onto V/K; and the pair (x), (y) is said to be a pair of corresponding generic points of U/K and V/K respectively under the rational transformation T, and we shall write $(y) = T(x)$. For every point (ξ) of U/K, its image under the rational transformation T consists of all the points (η) such that (ξ, η) is a specialization of (x, y) over the specialization $(x) \rightarrow (\xi)$ over K. If for a suitably chosen linear form $\sum_{j=0}^{t} a_j y_j$, the elements $y_i / \sum_{j=0}^{t} a_j y_j$, $i = 0, 1, \cdots, t$, are all in the quotient ring $Q(U/K, (\xi))$, then the transformation T is said to be regular at the point (ξ). In this case the image of (ξ) under T is a uniquely determined point of V/K. The rational transformation T is called birational, if conversely the point (x) is also rational over $K((y))$, i. e. if we have $K((x)) = K((y))$; the resulting rational transformation of V/K onto U/K is called the inverse T^{-1} of the transformation T. If a birational transformation T is regular at a point (ξ) and the inverse transformation T^{-1} is also regular at the image point (η) of (ξ) under T, then the transformation is said to be biregular at this point (ξ) or at the pair of points (ξ), (η). Two varieties U/K and V/K which are in biregular correspondence at the two points (ξ) and (η) respectively, have evidently the same quotient rings $Q(U/K, (\xi))$ and $Q(V/K, (\eta))$; hence they are also analytically equivalent at these two points. These concepts can be carried over to the case of two absolute varieties, just as we have done above for the definitions of quotient ring and analytical equivalence; we have then to replace the points (ξ) and (η) by two absolute subvarieties W and W' in U and V respectively, and the field K must be a field of definition for all the varieties U, V, W, W'.

2. Statement of the problem and result. Let U/K be a variety of r dimensions in the projective space S_m. As we have indicated at the beginning of the previous section, any positive d-cycle of degree n in U/K can be represented through its associated form by a point in the projective space S_t, and the set of all such points in S_t constitutes exactly the points of a bunch of varieties over K. Let V/K be any variety contained in this bunch, then the set of all the positive cycles in U/K corresponding to the points in V/K is called an algebraic system of positive cycles in U/K, and the variety V/K is called the associated variety of the algebraic system. Let (y) be a generic point of V/K, then the cycle $G(y)$ determined by (y) is called a generic cycle of the algebraic system and we shall denote the algebraic system itself by $|G(y)|$. We shall assume that the generic cycle $G(y)$ is prime rational over the field $K((y))$, and that a generic point (x) of $G(y)$ over $K((y))$ is also a generic point of U/K. This last assumption means simply that U/K is the carrier variety of the algebraic system $|G(y)|$. The pair of points (x), (y) then determines an algebraic correspondence T between the varieties U/K and V/K, which we shall call the associated correspondence of the algebraic system $|G(y)|$.

The problem with which we are concerned in this paper can be described as follows. Let $\bar{T}$ be any (K-irreducible) algebraic correspondence between the variety U/K and another variety $\bar{V}/K$ (which may of course lie in a different space). If $(\bar{y})$ is a generic point of $\bar{V}/K$, then the set of all points in U/K which correspond to $(\bar{y})$ under the correspondence $\bar{T}$ constitutes a variety or prime rational cycle $G(\bar{y})$ in U/K over the field $K((\bar{y}))$. If this cycle $G(\bar{y})$ has the dimension d and the degree n, then there corresponds to $G(\bar{y})$ by means of its natural coordinates a point (y) in the projective space S_t; let V/K be the variety in S_t determined by the generic point (y) over the field K. This variety V/K then determines an algebraic system of positive cycles in U/K, of which the cycle $G(y) = G(\bar{y})$ is a generic member, and there is an algebraic correspondence T between U/K and V/K associated with this algebraic system. Thus every algebraic correspondence $\bar{T}$ between U/K and a variety $\bar{V}/K$ induces in U/K an algebraic system of positive cycles, but in general the variety $\bar{V}/K$ is of course different from the associated variety V/K of the algebraic system and the correspondence $\bar{T}$ is different from the associated correspondence T. In fact, under the correspondence $\bar{T}$ between U/K and $\bar{V}/K$, the points of $\bar{V}/K$ will not in general be in one-to-one correspondence with the positive cycles of the algebraic system without exception; that is, to some points of $\bar{V}/K$ there might correspond more than one or even an infinite number of cycles of the algebraic system.[15] The question arises as to whether there are any other properties which distinguish the associated variety V/K from the rather arbitrarily chosen variety $\bar{V}/K$, apart from the fact that V/K represents in a one-to-one manner the cycles of the algebraic system $|G(y)|$. One would naturally ask for those properties which are invariant under a biregular birational transformation of the associated variety V/K. In terms of local geometry we can formulate the question as follows: Let $(\bar{\eta})$ be a point of $\bar{V}/K$ to which there corresponds a uniquely determined cycle $G(\bar{\eta})$ in U/K and let (η) be the corresponding point in V/K so that we have $G(\eta) = G(\bar{\eta})$. Then both varieties V/K and $\bar{V}/K$ can be said to represent through the correspondences T and $\bar{T}$ respectively the algebraic system of positive cycles $|G(y)| = |G(\bar{y})|$ in a one-to-one manner in the neighborhoods of the points (η) and $(\bar{\eta})$ respectively. The question is whether there are any locally birational or analytic properties which distinguish the associated correspondence T in the neighborhood of the pair (η), $G(\eta)$ from the correspondence $\bar{T}$ in the neighborhood of the pair $(\bar{\eta})$, $G(\bar{\eta})$. In this paper we shall give an answer to this question in the special case where the correspondence $\bar{T}$ is a rational transformation of U/K onto $\bar{V}/K$ and consequently the associated correspondence T is also a rational transformation of U/K onto V/K. We shall show that in this case the local analytic properties of the variety V/K in the neighborhood of the point (η) has a very simple connection with the local properties of the variety U/K in the neighborhood of the cycle $G(\eta)$, which is not true in general for the variety $\bar{V}/K$ in the neighborhood of $(\bar{\eta})$.

---

[15] For a more detailed description of the relations between the correspondences T and $\bar{T}$, see Chow and van der Waerden [5], § 2.

We shall from now on restrict ourselves to the case where the associated correspondence T of the algebraic system $|G(y)|$ is a rational transformation of U/K onto V/K. It is clear in this case that the variety V/K has the dimension $r - d$. We shall call such an algebraic system of positive cycles an involutional system, on account of the fact that in case $d = 0$ such a system is essentially what is usually called an involution in the classical algebraic geometry. For the sake of convenience, we shall assume that the ground field K contains infinitely many elements. Our main result is the following theorem:

THEOREM. *Let U/K be a variety; let $|G(y)|$ be an involutional system of positive cycles in U/K with the associated variety V/K and the associated transformation T, where (y) is a generic point of V/K. Let (η) be a rational point of V/K and (ξ) be a generic point of a K-component G'/K of $G(\eta)$; we assume that the field $K((\xi))$ is analytically disjoint*[16] *with respect to the quotient ring $Q(V/K, (\eta))$ over K; this implies in particular that*

---

[16] For the definition of analytical disjointness, see Appendix.

$K((y))$ and $K((\xi))$ are linearly disjoint over K, so that the product variety W/K of V/K and G'/K is defined. If the transformation T is regular at the point (ξ) and the variety G'/K is a simple K-component of $G(\eta)$ and is separably generated, then the varieties U/K and W/K are analytically equivalent at the points (ξ) and (η, ξ) respectively.

For the sake of greater generality, we have stated our theorem above for the "relative" varieties; this is the reason why we have to make the assumptions that the point (η) is rational and the field $K((\xi))$ is analytically disjoint with respect to the quotient ring $Q(V/K, (\eta))$ over K. In the applications, however, we are mostly concerned with the case where both varieties U/K and V/K are absolutely irreducible; in other words, we are concerned with an involutional system of positive cycles in an absolute variety U with an absolute variety V as the associated variety. It is then natural to assume that G' is an absolute variety, i. e. an (absolute) component of the positive cycle $G(\eta)$, and that the field K is the smallest field over which all the varieties U, V, G', (η) are defined. Since the field $K((\xi))$ is in this case a regular extension of K, it follows (Appendix, Theorem 4) that $K((\xi))$ is analytically disjoint with respect to any local domain with K as a basic field, hence in particular with respect to the local domain $Q(V/K, (\eta))$ over K. Recalling our definition of the analytical equivalence between two absolute varieties, we can state our theorem for absolute varieties as follows:

THEOREM. *Let U be an absolute variety; let $|G(y)|$ be an involutional system of positive cycles in U with the associated (absolute) variety V and the associated transformation T. Let (η) be any point of V and let G' be a component of $G(\eta)$. If G' is a simple component of $G(\eta)$ and the transformation T is regular at the subvariety G', then the variety U and the product variety $W = V \times G'$ are analytically equivalent at the subvarieties G' and $(\eta) \times G'$ respectively.*

From this theorem we can derive as an immediate corollary a criterion for simple points or subvarieties. Since the product variety W is simple at the subvariety $(\eta) \times G'$ if and only if both V and G' are simple at the subvarieties (η) and G' respectively, and since the variety G' is certainly simple at the variety G' itself, i. e. at its generic point, it follows that under the conditions of the above theorem the variety U is simple at the subvariety G' if and only if the variety V is simple at the point (η). It is evident that the application of this criterion to the problem of Jacobian variety gives an affirmative answer to the question we have posed at the beginning of the previous section.

Finally, we should like to add that the significance of our theorem goes beyond a mere criterion for simple points; it gives us a certain information about the analytic structure of U in the neighborhood of a cycle of an algebraic system. In order to express this more clearly, we introduce yet another definition. An involutional system $|G(y)|$ in an absolute variety U is called a fiber system if the transformation T is regular at every point of U and the cycle $G(\eta)$ is an absolute variety (i. e. consisting of one component with multiplicity 1) for every point (η) of V. It is evident that in this case each point of U belongs to exactly one cycle of the system, so that the entire variety U is fibered by the algebraic system of cycles. Our result can then be expressed roughly as follows: If $|G(y)|$ is a fiber system in U with the associated variety V and the associated transformation T, then U has properties similar to those of a fiber space in the topology, with V as the base space and T as the projection. In fact, there seems to be more than a formal analogy between this algebraic "fiber space" and the fiber space in topology; we hope to be able to come back to these questions on some future occasions.

Before we proceed with the proof of our theorem, we shall make a change in our terminology. It is clear that our main theorem is of a purely local nature, and as such it can be stated just as well for varieties in affine spaces instead of projective spaces. In fact, this is the form in which we shall prove the theorem in the next two sections; for, as is usually the case with a problem of purely local nature, it is very convenient to be able to work with the affine coordinates. It is easily seen that most of the definitions and concepts as developed and indicated in this and the previous sections can be carried over with suitable modifications to the varieties and points in affine spaces, and we can replace such field extensions as $K((x))$ by the ordinary field extension $K(x)$. Thus we shall from now on consider only varieties and points in a suitably chosen affine space in the projective space S_m; we shall denote this affine space also by S_m and represent a point in it by the affine coordinates $(x) = (x_1, \cdots, x_m)$. We shall assume that the affine S_m has been so chosen that the d-cycle $G(\eta)$ is finite; this implies of course that the variety U/K as well as the generic cycle $G(y)$ of the system are also finite. A variety is said to be finite with respect to an affine space if every generic point of it is finite (i. e. in the affine space), and a cycle is said to be finite if all its components are finite. Similarly, we can assume that the affine S_t has been so chosen that the point (η) is finite, which implies that the variety V/K is also finite; but for the sake of convenience later in our proof, we shall make a more definite choice of this affine space S_t with respect to a given choice of the affine coordinate system in the affine space S_m.

We set $\omega_0 = (Z_0 Z^{(1)}{}_1 \cdots Z^{(d)}{}_d)^n$ in the ordered set of monomials $\omega_0, \omega_1, \cdots, \omega_t$ in the $d+1$ set of indeterminates $(Z), (Z^{(1)}), \cdots, (Z^{(d)})$ as defined in the beginning of the previous section. Then the affine S_t is defined as the set of all points $(c_0, c_1, \cdots, c_t)$ with $c_0 \neq 0$, and such a point will be denoted by its affine coordinates $(c_1/c_0 \cdots, c_t/c_0)$. In order to insure that the point (η) is finite with respect to such a choice of the affine S_t, we shall assume that the affine coordinate system in the affine space S_m has been so chosen that the linear variety defined by the d equations $X_i = v_i$, $i = 1, \cdots, d$, where the $v_1, \cdots, v_d$ are independent variables over K, intersects the d-cycle $G(\eta)$ properly and finitely, i.e. the intersection consists of a finite number of finite points. Then the coefficient of the term ω_0 in the associated form $F_{(\eta)}(Z, Z^{(1)}, \cdots, Z^{(d)})$ of $G(\eta)$ cannot be zero. For, let $v^{(i)}{}_0 = -v_i$, $v^{(i)}{}_i = 1$, $v^{(i)}{}_j = 0$, for $i = 1, \cdots, d$, and $j \neq i$, then the form $F_{(\eta)}(Z, v^{(1)}, \cdots, v^{(d)})$ is the associated form of the intersection of $G(\eta)$ with the linear variety. It is easily seen that the coefficient of the term $Z_0{}^n$ in the form $F_{(\eta)}(Z, v^{(1)}, \cdots, v^{(d)})$ is exactly the coefficient of the term ω_0 in the form $F_{(\eta)}(Z, Z^{(1)}, \cdots, Z^{(d)})$. If this coefficient were zero, then the form $F_{(\eta)}(Z, v^{(1)}, \cdots, v^{(d)})$ must contain as factor at least one linear form without the term Z_0, and this linear form would then correspond to an infinite intersection point of $G(\eta)$ with the linear variety, in contradiction to our assumption. Thus the coefficient of the term ω_0 in the form $F_{(\eta)}(Z, Z^{(1)}, \cdots, Z^{(d)})$ is not zero, and this means that the point (η) is in the affine space S_t. From now on we shall assume that the associated form has been so normalized that the coefficient of the term ω_0 is equal to 1; the coefficients of the rest of the terms will then be the affine coordinates of the corresponding cycle.

3. Two lemmas.

We begin by recalling some well known facts concerning the semi-local rings. A ring $\mathfrak{o}$ is called a semi-local ring,[17] if it is a Noetherian ring and contains only a finite number of maximal prime ideals. We shall consider $\mathfrak{o}$ also as a topological space in the following way. Since $\mathfrak{o}$ is in particular a group with respect to the addition, we can define a topology in $\mathfrak{o}$ by specifying a fundamental system of neighborhoods of the element 0. Let $\mathfrak{a}$ be the product of all the maximal prime ideals of $\mathfrak{o}$, then the sets $\mathfrak{a}^s$, for $s = 1, 2, \cdots$, will be taken as a fundamental system of neighborhoods of 0. It can be shown that the usual conditions are satisfied; they are all trivial except the one asserting that $\bigcap_{s=1}^{\infty} \mathfrak{a}^s = 0$, which is the well known theorem due to Krull. Thus $\mathfrak{o}$ is a topological group and consequently a regular space; moreover, not only the addition but also the multiplication is a continuous function in $\mathfrak{o}$.

---

[17] Chevalley [1].

Let $\mathfrak{o}'$ be another semi-local ring which contains $\mathfrak{o}$ as a sub-ring, and let $\mathfrak{o}'$ be topologized in a similar way by means of the product $\mathfrak{a}'$ of its maximal prime ideals. The question arises as to the relation between the topologies of the two rings $\mathfrak{o}$ and $\mathfrak{o}'$. In general, they are not concordant; in other words, the sets $\mathfrak{a}'^s \cap \mathfrak{o}$ do not in general constitute a fundamental system of neighborhoods of 0 in $\mathfrak{o}$. However, in the important special case when $\mathfrak{o}'$ is a finite $\mathfrak{o}$-module (and this implies itself that $\mathfrak{o}'$ is a semi-local ring) and no non-zero element of $\mathfrak{o}$ is a zero-divisor in $\mathfrak{o}'$, it can be shown[18] that the sets $\mathfrak{a}'^s \cap \mathfrak{o}$ also constitute a fundamental system of neighborhoods of 0 in $\mathfrak{o}$, so that $\mathfrak{o}'$ contains $\mathfrak{o}$ not only as a sub-ring, but also as a sub-space. Explicitly this means that, in this case, there exist two sequences of positive integers $l(s)$ and $m(s)$, with $l(s) \rightarrow \infty$ and $m(s) \rightarrow \infty$ as $s \rightarrow \infty$, such that the relations $\mathfrak{a}^{l(s)} \subset \mathfrak{a}'^s \cap \mathfrak{o} \subset \mathfrak{a}^{m(s)}$ hold for all s.

Let U/K and V/K be two varieties of the same dimension r in the affine spaces S_m and S_t respectively. Let T be a rational transformation of U/K onto V/K, and let (x) and $(y) = T(x)$ be a pair of corresponding generic points of U/K and V/K respectively. Then $K(x)$ is a finite algebraic extension of $K(y)$; let n be the degree of the extension. To the generic point (y) of V/K there correspond in the inverse transformation T^{-1} n (not necessarily distinct) points $(x^{(1)}) = (x), (x^{(2)}), \cdots, (x^{(n)})$ of U/K, which are all generic points of U/K and which together constitute a complete set of conjugates over $K(y)$. We shall say that the inverse transformation T^{-1} is defined at a point (η) of V/K, if there is a finite specialization $(x^{(1)}, \cdots, x^{(n)}) \rightarrow (\xi^{(1)}, \cdots, \xi^{(n)})$ over the specialization $(y) \rightarrow (\eta)$, such that every specialization of $(x^{(1)}, \cdots, x^{(n)})$ over the specialization $(y) \rightarrow (\eta)$ coincides with this one except possibly for a reordering of the points $(\xi^{(1)}), \cdots, (\xi^{(n)})$. It is clear that if the rational transformation T is one which is associated with an involutional system of positive 0-cycles on U/K, so that V/K is the associated variety of the system, then the inverse transformation T^{-1} is defined at every point of the variety V/K, for a suitable choice of the affine spaces S_m and S_t.

LEMMA I. *Let T be a rational transformation of a variety U/K onto a variety V/K, both of the same dimension r. Let (ξ) be a point of U/K*

---

[18] See Chevalley [1], p. 699, Proposition 7. For the properties of semi-local rings used in the following, we refer once for all to this paper of Chevalley.

such that T *is regular at* (ξ) *and the inverse transformation* T^{-1} *is defined at the image point* $(\eta) = T(\xi)$. *Then the quotient ring* $\mathfrak{U} = Q(U/K, (\xi))$ *contains the quotient ring* $\mathfrak{V} = Q(V/K, (\eta))$ *as a subspace.*

Proof. Let $\mathfrak{r} = K[y]$ be the coordinate ring of V/K, and let $\mathfrak{p}$ be the prime ideal in $\mathfrak{r}$ determined by the point (η), and let $\mathfrak{s} = \mathfrak{r} - \mathfrak{p}$ be the multiplicatively closed system of all the elements of $\mathfrak{r}$ not in $\mathfrak{p}$. Then the quotient ring of $\mathfrak{r}$ with respect to $\mathfrak{s}$ is a local ring $\mathfrak{o}$, and the ideal $\mathfrak{m} = \mathfrak{o}\mathfrak{p}$ is the maximal prime ideal of $\mathfrak{o}$. It is easily seen that $\mathfrak{o}$ is the quotient ring $\mathfrak{V} = Q(V/K, (\eta))$ of V/K at (η). Let $\mathfrak{r}' = K[x]$ be the coordinate ring of U/K, and let $\mathfrak{p}'_1, \cdots, \mathfrak{p}'_n$ be the n prime ideals (not necessarily distinct) in $\mathfrak{r}'$ determined by the n points $(\xi^{(1)}), \cdots, (\xi^{(n)})$, and let $\mathfrak{s}' = \bigcap_{i=1}^{n} (\mathfrak{r}' - \mathfrak{p}'_i)$ be the multiplicatively closed system of all elements of $\mathfrak{r}'$ not in any one of the ideals $\mathfrak{p}'_1, \cdots, \mathfrak{p}'_n$. Let $\mathfrak{o}'$ be the quotient ring of $\mathfrak{r}'$ with respect to $\mathfrak{s}'$. Then $\mathfrak{o}'$ is a semi-local ring, and its maximal prime ideals are the ideals $\mathfrak{m}'_i = \mathfrak{o}'\mathfrak{p}'_i$ $(i = 1, \cdots, n)$, which are not necessarily distinct. Let $\mathfrak{o}'_i$ be the quotient ring of $\mathfrak{o}'$ with respect to the multiplicatively closed system $\mathfrak{o}' - \mathfrak{m}'_i$; each ring $\mathfrak{o}'_i$ is a local ring with the maximal prime ideal $\mathfrak{o}'_i\mathfrak{m}'_i$. It is clear that $\mathfrak{o}'_1$ is the quotient ring $\mathfrak{U} = Q(U/K, (\xi))$ of U/K at (ξ). We have to show that if $\mathfrak{o}'_1 \supset \mathfrak{o}$, then $\mathfrak{o}$ is a sub-space of $\mathfrak{o}'_1$. We shall achieve this by constructing a bigger ring which contains both $\mathfrak{o}'_1$ and $\mathfrak{o}$ as sub-spaces.

Consider the ring $\mathfrak{R} = K[x^{(1)}, \cdots, x^{(n)}]$, and for each i $(i = 1, \cdots, n)$, let $\mathfrak{P}_{ij}$, $j = 1, \cdots, a$, be the distinct prime ideals determined by the specializations of $(x^{(1)}, \cdots, x^{(n)})$ over the specialization $(x^{(i)}) \rightarrow (\xi^{(1)})$ over K. Since there is an automorphism of the field $k(x^{(1)}, \cdots, x^{(n)})$ over $K(y)$ which carries any one of the points $(x^{(i)})$ into any other, it follows that we can arrange the ideals $\mathfrak{P}_{ij}$ in such a way that the ideals $\mathfrak{P}_{1j}, \cdots, \mathfrak{P}_{nj}$ are conjugate to each other for every $j = 1, 2, \cdots, a$. Let $\mathfrak{S} = \bigcap_{i,j} (\mathfrak{R} - \mathfrak{P}_{ij})$ be the multiplicatively closed system of all elements of $\mathfrak{R}$ not contained in any one of the prime ideals $\mathfrak{P}_{ij}$, and let $\mathfrak{O}$ be the quotient ring of $\mathfrak{R}$ with respect to $\mathfrak{S}$. The ring $\mathfrak{O}$ is then a semi-local ring, and its maximal prime ideals are the ideals $\mathfrak{M}_{ij} = \mathfrak{O}\mathfrak{P}_{ij}$. Let $\bar{\mathfrak{O}}$ be the integral closure of $\mathfrak{O}$; since $\bar{\mathfrak{O}}$ is a finite $\mathfrak{O}$-module, it is also a semi-local ring. For every i and j, let $\bar{\mathfrak{M}}_{ijk}$ $(k = 1, \cdots, b)$ be the distinct maximal prime ideals of $\bar{\mathfrak{O}}$ such that $\bar{\mathfrak{M}}_{ijk} \cap \mathfrak{O} = \mathfrak{M}_{ij}$; and we shall arrange these ideals in such a way that the ideals $\bar{\mathfrak{M}}_{1jk}, \cdots, \bar{\mathfrak{M}}_{njk}$ are conjugate to each other, for every j and k. Now, since every element of $\mathfrak{o}$ is finite over any finite specialization of the elements of $\mathfrak{O}$, it is integrally dependent on $\mathfrak{O}$; hence, we have $\bar{\mathfrak{O}} \supset \mathfrak{o}$. On the other hand, since every element of $\bar{\mathfrak{O}}$ is also finite over any finite specialization of the elements of $\mathfrak{o}$ and is consequently integrally dependent on $\mathfrak{o}$, we conclude that $\bar{\mathfrak{O}}$ is the integral closure of $\mathfrak{o}$ in the field $K(x^{(1)},\cdots,x^{(n)})$. From this it follows that $\bar{\mathfrak{O}}$ is a finite $\mathfrak{o}$-module, and hence $\bar{\mathfrak{O}}$ contains $\mathfrak{o}$ as a sub-space, and we have $\bar{\mathfrak{M}}_{ijk} \cap \mathfrak{o} = \mathfrak{m}$.

It is clear from our construction that $\mathfrak{R} \supset \mathfrak{r}'$; and, as all the prime ideals $\mathfrak{P}_{ij}$ of $\mathfrak{R}$ contract in $\mathfrak{r}'$ to one of ideals $\mathfrak{p}'_1,\cdots,\mathfrak{p}'_n$, we have $\mathfrak{r}' \cap \mathfrak{S} \subset \mathfrak{s}'$, from which it follows that $\mathfrak{O} \supset \mathfrak{o}'$. The ideals $\mathfrak{P}_{1j}$ $(j = 1,\cdots,a)$ are evidently the only maximal prime ideals of $\mathfrak{R}$ which contract to $\mathfrak{p}'_1$ in $\mathfrak{r}'$. It follows that the ideals $\mathfrak{M}_{1j}$ $(j = 1,\cdots,a)$ consist of all the maximal prime ideals of $\mathfrak{O}$ which contract to $\mathfrak{m}'_1$ in $\mathfrak{o}'$. Hence the ideals $\bar{\mathfrak{M}}_{1jk}$ are all the maximal prime ideals of $\bar{\mathfrak{O}}$ which contract to $\mathfrak{m}'_1$ to $\mathfrak{o}'$. Let $\bar{\mathfrak{O}}_1$ be the quotient ring of $\mathfrak{O}$ with respect to the multiplicatively closed system $\bigcap_{j,k} (\bar{\mathfrak{O}} - \bar{\mathfrak{M}}_{1jk})$. Since $\bar{\mathfrak{M}}_{1jk} \cap \mathfrak{o}' = \mathfrak{m}'_1$, the ring $\mathfrak{o}'_1$ is contained in $\bar{\mathfrak{O}}_1$. On the other hand, it is easily seen that every element of $\bar{\mathfrak{O}}_1$ is finite over any finite specialization of the elements of $\mathfrak{o}'_1$ and hence is integrally dependent on $\mathfrak{o}'_1$. Therefore, $\bar{\mathfrak{O}}_1$ is the integral closure of $\mathfrak{o}'_1$ in the field $K(x^{(1)},\cdots,x^{(n)})$, from which it follows that $\bar{\mathfrak{O}}_1$ is a finite $\mathfrak{o}'_1$-modul and consequently contains $\mathfrak{o}'_1$ as a subspace.

We shall now show that $\bar{\mathfrak{O}}_1$ contains also $\mathfrak{o}$ as a sub-space. Let $\mathfrak{A}_i = \prod_{j,k} \bar{\mathfrak{M}}_{ijk}$ and $\mathfrak{A} = \prod'_{i,j,k} \bar{\mathfrak{M}}_{jik}$, where the product $\prod'$ extends over all the distinct ideals only among the $\bar{\mathfrak{M}}_{ijk}$. Since $\bar{\mathfrak{O}}$ contains $\mathfrak{o}$ as a sub-space, and since the sets $\mathfrak{A}^s$ $(s = 1, 2,\cdots)$ constitute by definition a fundamental system of neighborhoods of 0 in $\bar{\mathfrak{O}}$, we have the relations $\mathfrak{a}^{l(s)} \subset \mathfrak{A}^s \cap \mathfrak{o} \subset \mathfrak{a}^{m(s)}$ with $l(s) \to \infty$ and $m(s) \to \infty$, as $s \to \infty$. Since any two distinct ones of the ideals $\bar{\mathfrak{M}}_{ijk}{}^s$ are relatively prime to each other, we have

$$\mathfrak{A}^s = \prod_{i,j,k}{}' \bar{\mathfrak{M}}_{ijk}{}^s = \bigcap_{i,j,k} \bar{\mathfrak{M}}_{ijk}{}^s = \bigcap_i \Big(\bigcap_{j,k} \bar{\mathfrak{M}}_{ijk}{}^s\Big) = \bigcap_i \Big(\prod_{j,k} \bar{\mathfrak{M}}_{ijk}{}^s\Big)$$
$$= \bigcap_i \mathfrak{A}_i{}^s.$$

Hence we have $\mathfrak{a}^{l(s)} \subset \bigcap_i (\mathfrak{A}_i{}^s \cap \mathfrak{o}) \subset \mathfrak{a}^{m(s)}$. Now, since the ideals $\mathfrak{A}_i{}^s$ are all conjugate to each other over $K(y)$, we have evidently $\mathfrak{A}_1{}^s \cap \mathfrak{o} = \cdots = \mathfrak{A}_n{}^s \cap \mathfrak{o}$. Hence we have the relations $\mathfrak{a}^{l(s)} \subset \mathfrak{A}_1{}^s \cap \mathfrak{o} \subset \mathfrak{a}^{m(s)}$, which means that the sets $\mathfrak{A}_1{}^s \cap \mathfrak{o}$ constitute a fundamental system of neighborhoods of 0 in $\mathfrak{o}$. Now, the sets $(\mathfrak{O}_1\mathfrak{A}_1)^s$ constitute by definition a fundamental system of neighborhoods of 0 in $\bar{\mathfrak{O}}_1$, and we have from the theory of quotient rings[19] that $(\bar{\mathfrak{O}}_1\mathfrak{A}_1)^s \cap \bar{\mathfrak{O}} = \mathfrak{A}_1{}^s$. Therefore we have $(\bar{\mathfrak{O}}_1\mathfrak{A}_1)^s \cap \mathfrak{o} = (\bar{\mathfrak{O}}_1\mathfrak{A}_1)^s \cap \bar{\mathfrak{O}} \cap \mathfrak{o} = \mathfrak{A}_1{}^s \cap \mathfrak{o}$, which shows that $\mathfrak{o}$ is a sub-space of $\bar{\mathfrak{O}}_1$.

---

[19] We observe that since the $\bar{\mathfrak{M}}_{1jk}$ are maximal prime ideals in $\mathfrak{O}$, we have $\mathfrak{A}_1{}^s = \bigcap_{j,k} \bar{\mathfrak{M}}_{1jk}{}^s$, and each $\bar{\mathfrak{M}}_{1jk}{}^s$ is a primary ideal. Therefore we have $\bar{\mathfrak{O}}_1\bar{\mathfrak{M}}_{1jk}{}^s \cap \bar{\mathfrak{O}} = \bar{\mathfrak{M}}_{1jk}{}^s$ and hence also $\bar{\mathfrak{O}}_1\mathfrak{A}_1{}^s \cap \bar{\mathfrak{O}} = \mathfrak{A}_1{}^s$.

Since the ring $\bar{\mathfrak{O}}_1$ contains both $\mathfrak{o}'_1$ and $\mathfrak{o}$ as sub-spaces, and since (on account of the regularity of the rational transformation T at $(\xi^{(1)})$) $\mathfrak{o}'_1$ contains $\mathfrak{o}$, it follows that $\mathfrak{o}'_1$ contains $\mathfrak{o}$ as a sub-space. Thus the lemma is proved.

[Note added in proof (March 10, 1950): Professor Zariski has kindly called our attention to the fact that our Lemma I is closely related to a result in his recent note in the *Proceedings of the National Academy of Sciences*, vol. 35 (1949), pp. 62-66, Theorem 3. In fact, if the variety V/K is analytically irreducible at the point (η), our Lemma I is an immediate consequence of Zariski's theorem. In case V/K is not analytically irreducible, it is also possible to deduce our Lemma I from Zariski's theorem, if we make use of Zariski's theory of normal varieties. This can be done in the following way, as communicated to us by Professor Zariski.

Let $\bar{U}/K$ and $\bar{V}/K$ be derived normal models of U/K and V/K respectively. Let $(\bar{\xi}^{(1)}), (\bar{\xi}^{(2)}), \cdots, (\bar{\xi}^{(g)})$ be the points of $\bar{U}/K$ which correspond to (ξ) and let similarly $(\bar{\eta}^{(1)}), (\bar{\eta}^{(2)}), \cdots, (\bar{\eta}^{(h)})$ be the points of $\bar{V}/K$ which correspond to (η). Since (η) is the only point of V/K which corresponds to (ξ) under T, the points of $\bar{V}/K$ which correspond to any $(\bar{\xi}^{(i)})$ (in the rational transformation $\bar{T}$ of $\bar{U}/K$ onto $\bar{V}/K$) are among the points $(\bar{\eta}^{(j)})$ and hence are finite in number. Since $\bar{U}/K$ is normal it follows that to each point $(\bar{\xi}^{(i)})$ there corresponds a unique point, say $(\bar{\eta}^{(\mu_i)})$, and that $\bar{T}$ is regular at $(\bar{\xi}^{(i)})$. Recalling the fact that a normal variety is analytically irreducible at any one of its points (see Zariski, *Annals of Mathematics*, vol. 49 (1948), pp. 352-361), it follows from Zariski's theorem that

(1) the quotient ring of $(\bar{\eta}^{(\mu_i)})$ is a sub-space of the quotient ring of $(\bar{\xi}^{(i)})$ $(i = 1, 2, \cdots, g)$.

Now let us assume that in every specialization $(x^{(1)}, x^{(2)}, \cdots, x^{(n)}) \to (\xi^{(1)}, \xi^{(2)}, \cdots, \xi^{(n)})$ over the specialization $(y) \to (\eta)$ the point (ξ) occurs at least once among the $(\xi^{(i)})$ (this assumption is weaker than the assumption that the inverse transformation T^{-1} is defined at (η)). We have for each $j = 1, 2, \cdots, h$ the specialization $(y, \bar{y}) \to (\eta, \bar{\eta}^{(j)})$ (where $(\bar{y})$ is the point $\bar{V}/K$ which corresponds to the general point (y) of V/K), and if we apply the above assumption to the specializations of $(x^{(1)}, x^{(2)}, \cdots, x^{(n)})$ which are over the specialization $(y, \bar{y}) \to (\eta, \bar{\eta}^{(j)})$ we conclude at once that

(2) each of the h points $(\bar{\eta}^{(j)})$ occurs at least once among the points $(\bar{\eta}^{(\mu_i)})$, $i = 1, 2, \cdots, g$.

From (1) and (2) it follows that the intersection of the quotient rings $\mathfrak{o}(\bar{\eta}^{(j)})$ of the h points $(\bar{\eta}^{(j)})$ is a sub-space of the intersection of the quotient rings $\mathfrak{o}'(\bar{\xi}^{(i)})$ of the g points $(\bar{\xi}^{(i)})$. Since $\bar{V}/K$ is a derived normal model of V/K, the quotient ring $\mathfrak{o}$ of the point (η) is a sub-space of $\bigcap_j \mathfrak{o}(\bar{\eta}^{(j)})$. *Hence* $\mathfrak{o}$ *is a sub-space of* $\bigcap_i \mathfrak{o}'(\bar{\xi}^{(i)})$. Since also $\bar{U}/K$ is a derived normal model of U/K, *the quotient ring* $\mathfrak{o}'_1$ *of the point* $(\xi) = (\xi^{(1)})$ *is a sub-space of* $\bigcap_i \mathfrak{o}'(\bar{\xi}^{(i)})$. From these conclusions and from the fact that the space $\mathfrak{o}$ is a *subset* of the space $\mathfrak{o}'_1$ (since T is regular at (ξ)), our Lemma I follows.]

Before we proceed to the second lemma, we shall make a few remarks about the coefficients of the associated form. Let U/K be a variety in S_m. Consider an involutional system of positive 0-cycles of degree n in U/K with the associated variety V/K and the associated transformation T, and let (x) and $y = T(x)$ be a pair of corresponding generic points of U/K and V/K respectively. Let $F_{(y)}(Z)$ be the associated form of the generic 0-cycle $G(y)$ of the system. We have then

$$F_{(y)}(Z) = \prod_{i=1}^{n} (Z_0 + \sum_{j=1}^{m} Z_j x^{(i)}{}_j),$$

where the n points $(x^{(1)}) = (x)$, $(x^{(2)}), \cdots, (x^{(n)})$ form a complete set of conjugates over $K(y)$. The coordinates of the point (y) are the coefficients of the form $F_{(y)}(Z)$, and they are by hypothesis rational functions of (x). Let $f_k(Z_0, Z_k)$ be the form obtained from $F_{(y)}(Z)$ by setting $Z_j = 0$ for all $j \neq 0, k$; we have then evidently

$$f_k(Z_0, Z_k) = \prod_{i=1}^{n} (Z_0 + Z_k x^{(i)}{}_k), \qquad k = 1, \cdots, m.$$

Thus, for each $k = 1, \cdots, m$, the coefficients of the form $f_k(Z_0, Z_k)$ are the elementary symmetric functions of the n elements $x^{(1)}{}_k, \cdots, x^{(n)}{}_k$. It is clear that the coefficients of the m forms $f_k(Z_0, Z_k)$, $k = 1, \cdots, m$, are all part of the coefficients of the form $F_{(y)}(Z)$; that is, they all occur among the coordinates of the point (y). Furthermore, let

$$g_k(Z_0, Z_k) = \prod_{i=2}^{n} (Z_0 + Z_k x^{(i)}{}_k), \qquad k = 1, \cdots, m$$

so that we have $f_k(Z_0, Z_k) = (Z_0 + Z_k x_k) g_k(Z_0, Z_k)$, $k = 1, \cdots, m$. It is well known that for each $k = 1, \cdots, m$, the coefficients of $g_k(Z_0, Z_k)$, being the elementary symmetric functions of the $n - 1$ elements $x^{(2)}{}_k, \cdots, x^{(n)}{}_k$, can be expressed as polynomials in the coefficients of the form $f_k(Z_0, Z_k)$ and the element x_k. It follows then that the coefficients of the forms $g_k(Z_0, Z_k)$, $k = 1, \cdots, m$, are all elements of the ring $K[x, y]$. We shall make use of this remark presently.

LEMMA II. *Let U/K be a variety; let $|\, G(y) \,|$ be an involutional system of positive 0-cycles in U/K with the associated variety V/K and the associated transformation T. Let (ξ) be a point of U/K such that the transformation T is regular at (ξ), and that the corresponding point $(\eta) = T(\xi)$ on V/K determines an 0-cycle $G(\eta)$ which contains (ξ) as a simple component. Then the maximal prime ideal $\mathfrak{u}$ of the quotient ring $\mathfrak{U} = Q(U/K, (\xi))$ has a basis consisting of elements in the quotient ring $\mathfrak{B} = Q(V/K, (\eta))$.*

Proof. Let (x) and $(y) = T(x)$ be a pair of corresponding generic points of U/K and V/K respectively. The hypothesis that the point (ξ) is a simple component of the special cycle $G(\eta)$ implies that the generic cycle $G(y)$ can only have simple components; hence the field $K(x)$ must be a separable algebraic extension of $K(y)$ and the n points $(x^{(1)}) = (x), (x^{(2)}), \cdots, (x^{(n)})$ of the generic cycle $G(y)$ are distinct. Let $(x^{(1)}, \cdots, x^{(n)}) \rightarrow (\xi^{(1)}, \cdots, \xi^{(n)})$ be any specialization over the specialization $(x) \rightarrow (\xi)$ over K; the points $(\xi^{(1)}), \cdots, (\xi^{(n)})$ are then evidently the n points of the cycle $G(\eta)$. Since the point (ξ) occurs only once in the cycle $G(\eta)$, we have $(\xi^{(i)}) \neq (\xi)$ for all $i \neq 1$. Therefore, we can assume that the affine coordinate system in S_m has been so chosen that we have $\xi^{(i)}{}_j \neq \xi_j$ for all $i \neq 1$ and all j. It is easily seen that this condition can always be obtained by means of a suitably chosen affine transformation with coefficients in K, the field K being assumed to contain infinitely many elements. Let $\lambda = (i_1, \cdots, i_m)$ be any ordered set of m positive integers not greater than n, we shall set $(\xi^\lambda) = (\xi^{(i_1)}{}_1, \cdots, \xi^{(i_m)}{}_m)$. Then, as λ runs through all the n^m possible ordered sets $(i_1, \cdots, i_m)$ of positive integers not greater than n, we obtain a set of n^m (not necessarily distinct) points $\{(\xi^\lambda)\}$. Since we have $\xi^{(i)}{}_j \neq \xi_j$ for all $i \neq 1$ and all j, it is clear that the point (ξ) occurs only once among the n^m points of the set (ξ^λ). Similarly, we set $(x^\lambda) = (x^{(i_1)}{}_1, \cdots, x^{(i_m)}{}_m)$ and obtain a set of n^m points $\{(x^\lambda)\}$. It is clear that for every λ the specialization $(x^\lambda) \rightarrow (\xi^\lambda)$ is the only specialization over the specialization $(x^{(1)}, \cdots, x^{(n)}) \rightarrow (\xi^{(1)}, \cdots, \xi^{(n)})$ over K.

Let $\psi_j(X)$ $(j = 1, \cdots, h)$ be a basis of the prime ideal determined by the point (ξ) in the polynomial ring $K[X] = K[X_1, \cdots, X_m]$. It is easily seen that the h elements $\psi_j(x)$, $j = 1, \cdots, h$, constitute then also a basis of the ideal $\mathfrak{u}$ in the local ring $\mathfrak{U}$. Since the point (ξ) occurs only once among the set $\{(\xi^\lambda)\}$, it follows that for every $\lambda \neq (1, \cdots, 1)$ the h elements

$\psi_j(\xi^\lambda)$, $j=1,\cdots,h$, are not all zero. There exist then h linearly independent combinations

$$\phi_j(X)=\sum_{k=1}^{h}a_{jk}\psi_k(X), \qquad j=1,\cdots,h,$$

of the polynomials $\psi_j(X)$ with coefficients a_{jk} in K, such that for every $\lambda\neq(1,\cdots,1)$ none of the h elements $\phi_j(\xi^\lambda)$, $j=1,\cdots,h$, vanishes. The elements $\phi_j(x)$, $j=1,\cdots,h$, constitute of course also a basis of the ideal $\mathfrak{u}$ in $\mathfrak{U}$; and we have for each λ and j the unique specialization $\phi_j(x^\lambda)\rightarrow\phi_j(\xi^\lambda)$ over the specialization $(x^{(1)},\cdots,x^{(n)})\rightarrow(\xi^{(1)},\cdots,\xi^{(n)})$ over K.

Now we set

$$\Phi_j=\prod_\lambda\phi_j(x^\lambda), \qquad j=1,\cdots,h,$$

where the product runs over all the n^m points of the set $\{(x^\lambda)\}$. We prove first that these h elements Φ_j, $j=1,\cdots,h$, are elements of the ring $K[y]$; this evidently implies that they are also elements of the quotient ring $\mathfrak{B}$. To show this, we set $\phi_j(X)=\phi_j^{(0)}(X_1,\cdots,X_m)$, $j=1,\cdots,h$; we can then write

$$\Phi_j=\prod_{i_m=1}^{n}\cdots\prod_{i_1=1}^{n}\phi_j^{(0)}(x^{(i_1)}{}_1,\cdots,x^{(i_m)}{}_m), \qquad j=1,\cdots,h.$$

Consider the polynomials

$$\phi_j^{(1)}(X_2,\cdots,X_m)=\prod_{i_1=1}^{n}\phi_j^{(0)}(x^{(i_1)}{}_1,X_2,\cdots,X_m), \qquad j=1,\cdots,h;$$

we have evidently

$$\Phi_j=\prod_{i_m=1}^{n}\cdots\prod_{i_2=1}^{n}\phi_j^{(1)}(x^{(i_2)}{}_2,\cdots,x^{(i_m)}{}_m), \qquad j=1,\cdots,h.$$

Since the coefficients of the polynomials $\phi_j^{(1)}(X_2,\cdots,X_m)$, $j=1,\cdots,h$, are integral rational symmetric functions of the n elements $x^{(1)}{}_2,\cdots,x^{(n)}{}_2$ they can be expressed as integral rational functions of the elementary symmetric functions of these n elements. This means that the coefficients of the polynomials $\phi_j^{(1)}(X_2,\cdots,X_m)$, $j=1,\cdots,h$, are integral rational functions of the coefficients of the form $f_1(Z_0,Z_1)$; and since the coefficients of $f_1(Z_0,Z_1)$ are themselves some of the coordinates of the point (y), it follows that the coefficients of $\phi_j^{(1)}(X_2,\cdots,X_m)$, $j=1,\cdots,h$, are elements of $K[y]$. Consider next the polynomials

$$\phi_j^{(2)}(X_3,\cdots,X_m)=\prod_{i_2=1}^{n}\phi_j^{(1)}(x^{(i_2)}{}_2,X_3,\cdots,X_m), \qquad j=1,\cdots,h;$$

we have evidently

$$\Phi_j = \prod_{i_m=1}^{n} \cdot \cdot \cdot , \prod_{i_3=1}^{n} \phi_j{}^{(2)} (x^{(i_3)}{}_3, \cdot \cdot \cdot , x^{(i_m)}{}_m), \qquad j = 1, \cdot \cdot \cdot , h.$$

Since the coefficients of the polynomials $\phi_j{}^{(2)}(X_3, \cdot \cdot \cdot, X_m)$, $j = 1, \cdot \cdot \cdot, h$, are integral rational symmetric functions of the n elements $x^{(1)}{}_2, \cdot \cdot \cdot, x^{(n)}{}_2$ with coefficients in $K[y]$, it follows that they can be expressed as integral rational functions of the elementary symmetric functions of these n elements also with coefficients in $K[y]$. This means that the coefficients of the polynomials $\phi_j{}^{(2)}(X_3, \cdot \cdot \cdot, X_m)$, $j = 1, \cdot \cdot \cdot, h$, are integral rational functions of the coefficients of the form $f_2(Z_0, Z_2)$, with coefficients in $K[y]$; and since the coefficients of the form $f_2(Z_0, Z_2)$ are themselves some of the coordinates of the point (y), it follows that the coefficients of $\phi_j{}^{(2)}(X_3, \cdot \cdot \cdot, X_m)$, $j = 1, \cdot \cdot \cdot, h$, are elements of $K[y]$. Thus, proceeding successively in this manner, we conclude after m steps that the elements Φ_j, $j = 1, \cdot \cdot \cdot, h$, are elements of $K[y]$.

We maintain that the h elements Φ_j, $j = 1, \cdot \cdot \cdot, h$, constitute a basis of the ideal $\mathfrak{u}$ in the ring $\mathfrak{U}$. Since we have $\Phi_j = (\Phi_j/\phi_j(x))\phi_j(x)$, $j = 1, \cdot \cdot \cdot, h$, and since the h elements $\phi_j(x)$, $j = 1, \cdot \cdot \cdot, h$, constitute a basis of $\mathfrak{u}$ in $\mathfrak{U}$, it is sufficient to prove that the elements $\Phi_j/\phi_j(x)$, $j = 1, \cdot \cdot \cdot, h$, are units of the ring $\mathfrak{U}$. We first show that these elements are elements of the ring $\mathfrak{U}$. In fact, indicating by $\prod_{(k)}^{(m)}{}'$ the product over all the $m - k + 1$ indices $i_k, \cdot \cdot \cdot, i_m$ with the exception of the one combination $i_k = \cdot \cdot \cdot = i_m = 1$, we obtain by a simple calculation the following:

$$\begin{aligned}
\Phi_j/\phi_j(x) &= \prod_{(1)}^{(m)}{}' \phi_j{}^{(0)} (x^{(i_1)}{}_1, \cdot \cdot \cdot , x^{(i_m)}{}_m) \\
&= \prod_{i_1=2}^{n} \phi_j{}^{(0)} (x^{(i_1)}{}_1, x_2, \cdot \cdot \cdot , x_m) \prod_{(2)}^{(m)}{}' \phi_j{}^{(1)} (x^{(i_2)}{}_2, \cdot \cdot \cdot , x^{(i_m)}{}_m) \\
&= \cdot \cdot \cdot \\
&= \prod_{k=1}^{m} \Big(\prod_{i_k=2}^{n} \phi_j{}^{(k-1)} (x^{(i_k)}{}_k, x_{k+1}, \cdot \cdot \cdot , x_m) = \prod_{k=1}^{m} \theta_j{}^{(k)} (x_{k+1}, \cdot \cdot \cdot , x_m), \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (j = 1, \cdot \cdot \cdot , h),
\end{aligned}$$

where in the last step we have set

$$\theta_j{}^{(k)} (X_{k+1}, \cdot \cdot \cdot , X_m) = \prod_{i_k=2}^{n} \phi_j{}^{(k-1)} (x^{(i_k)}{}_k, X_{k+1}, \cdot \cdot \cdot , X_m); \qquad k = 1, \cdot \cdot \cdot , m; j = 1, \cdot \cdot \cdot , h.$$

Thus it is only necessary to prove that for each $k = 1, \cdot \cdot \cdot, m$ and each $j = 1, \cdot \cdot \cdot, h$, the element $\theta_j{}^{(k)}(x_{k+1}, \cdot \cdot \cdot, x_m)$ is an element of $\mathfrak{U}$. Since the coefficients of the polynomial $\theta_j{}^{(k)}(X_{k+1}, \cdot \cdot \cdot, X_m)$ are integral rational symmetric functions of the $n - 1$ elements $x^{(2)}{}_k, \cdot \cdot \cdot, x^{(n)}{}_k$ with coefficients in $K[y]$, they can be expressed as integral rational functions of the elementary symmetric functions of these $n-1$ elements, also with coefficients in $K[y]$. This means that the coefficients of $\theta_j^{(k)}(X_{k+1},\cdots,X_m)$ are integral rational functions of the coefficients of the form $g_k(Z_0,Z_k)$, with coefficients in $K[y]$; and since the coefficients of $g_k(Z_0,Z_k)$ are themselves elements of $K[x,y]$, it follows that the coefficients of $\theta_j^{(k)}(X_{k+1},\cdots,X_m)$ are also elements of $K[x,y]$. Therefore, the element $\theta_j^{(k)}(x_{k+1},\cdots,x_m)$ is also an element of $K[x,y]$; and since the transformation T is regular at the point (ξ), the element $\theta_j^{(k)}(x_{k+1},\cdots,x_m)$ is in the quotient ring $\mathfrak{U}$. Thus we have shown that the elements $\Phi_j/\phi_j(x)$, $j=1,\cdots,h$, are all elements of the ring $\mathfrak{U}$.

Finally, we show that the elements $\Phi_j/\phi_j(x)$, $j=1,\cdots,h$, are units of the ring $\mathfrak{U}$. In fact, we have

$$\Phi_j/\phi_j(x)=\prod_\lambda{}'\,\phi_j(x^\lambda), \qquad j=1,\cdots,h,$$

where the product $\prod'_\lambda$ runs over all the n^m points of the set $\{(x^\lambda)\}$ except the point (x). For every λ and j, we have the unique specialization $\phi_j(x^\lambda)\to\phi_j(\xi^\lambda)$ over the specialization $(x^{(1)},\cdots,x^{(n)})\to(\xi^{(1)},\cdots,\xi^{(n)})$ over K; hence we have also the unique specialization

$$\Phi_j/\phi_j(x)=\prod_\lambda{}'\,\phi_j(x^\lambda)\to\prod_\lambda{}'\,\phi_j(\xi^\lambda), \qquad j=1,\cdots,h,$$

over the specialization $(x^{(1)},\cdots,x^{(n)})\to(\xi^{(1)},\cdots,\xi^{(n)})$ over K. Since the elements $\Phi_j/\phi_j(x)$, $j=1,\cdots,h$, are elements of $\mathfrak{U}$, it follows that the specialization

$$\Phi_j/\phi_j(x)\to\prod_\lambda{}'\,\phi_j(\xi^\lambda), \qquad j=1,\cdots,h,$$

is the uniquely determined specialization of these h elements over the specialization $(x)\to(\xi)$ over K. Now, we have shown above that $\phi_j(\xi^\lambda)\neq 0$ for every $\lambda\neq(1,\cdots,1)$ and every j; hence we have also

$$\prod_\lambda{}'\,\phi_j(\xi^\lambda)\neq 0, \qquad j=1,\cdots,h.$$

This means that the h elements $\Phi_j/\phi_j(x)$, $j=1,\cdots,h$, are not in $\mathfrak{u}$ and hence units of $\mathfrak{U}$. This concludes the proof of Lemma II.

4. Proof of the theorem. We restate here our theorem with slight modifications for affine varieties:

THEOREM. *Let U/K be a variety; let $|G(y)|$ be an involutional system of positive cycles in U/K with the associated variety V/K and the associated transformation T, where (y) is a generic point of V/K. Let (η) be a* *rational point of V/K such that the cycle $G(\eta)$ is finite, and let (ξ) be a generic point of a K-component G'/K of $G(\eta)$. We assume that the field $K(\xi)$ is analytically disjoint with respect to the ring $Q(V/K, (\eta))$ over K; this implies in particular that the fields $K(\xi)$ and $K(y)$ are linearly disjoint over K, so that the product variety $W/K = V/K \times G'/K$ is defined. If the transformation T is regular at the point (ξ) and if the variety G'/K is a simple K-component of $G(\eta)$ and is separably generated, then the varieties U/K and W/K are analytically equivalent at the points (ξ) and (η, ξ) respectively.*

Proof. As before, let (x) and $(y) = T(x)$ be a pair of corresponding generic points of U/K and V/K respectively. We shall first prove the theorem for the case when the dimension d of the positive cycles of the involutional system is zero and then extend it by a well known device to the general case.

(a) The case $d = 0$. The assumption that G'/K is a simple K-component of $G(\eta)$ and is separably generated implies that the point (ξ) is a simple component of the 0-cycle $G(\eta)$; hence the hypotheses of both Lemma I and Lemma II are satisfied. Let $\mathfrak{U} = Q(U/K, (\xi))$ be the quotient ring of U/K at the point (ξ) and $\mathfrak{u}$ be the maximal prime ideal in $\mathfrak{U}$; let $\mathfrak{V} = Q(V/K, (\eta))$ be the quotient ring of V/K at the point (η); and let $\mathfrak{W} = (W/K, (\eta, \xi))$ be the quotient ring of W/K at the point (η, ξ). Let $\mathfrak{U}^*$ be the completion of the local ring $\mathfrak{U}$ and let $\mathfrak{u}^*$ be its maximal prime ideal; let $\mathfrak{V}^*$ be the completion of the local ring $\mathfrak{V}$ and $\mathfrak{v}^*$ be its maximal prime ideal; let $\mathfrak{W}^*$ be the completion of the local ring $\mathfrak{W}$. Our aim is to prove that the two rings $\mathfrak{U}^*$ and $\mathfrak{W}^*$ are K-isomorphic. We observe first that $\mathfrak{W}$ is also the quotient ring $Q(V/K(\xi), (\eta))$ of the extension variety $V/K(\xi)$ of V/K over $K(\xi)$ at the point (η); which means that $\mathfrak{W}$ is the extension of $\mathfrak{V}$ over the field $K(\xi)$. Furthermore, since $K(\xi)$ is a finite extension of K and since $\mathfrak{V}$ contains K as a coefficient field, it follows that $\mathfrak{W}^* = K(\xi) \times \mathfrak{V}^*$. According to Lemma I, $\mathfrak{U}$ contains $\mathfrak{V}$ as a sub-space; hence $\mathfrak{U}^*$ contains $\mathfrak{V}^*$ as a subring. Since $K(\xi)$ is separable over K, $\mathfrak{U}$ contains K as a basic field; hence the completion $\mathfrak{U}^*$ contains a coefficient field K' over K, which is K-isomorphic to the residue ring $\mathfrak{U}^*/\mathfrak{u}^*$ and hence also K-isomorphic to $K(\xi)$. The field K' is not in general a subfield of the universal domain $\mathfrak{K}$ and is hence an "abstract" field in our terminology; however, since K' is K-isomorphic with the field $K(\xi)$, we can without loss of generality identify the two fields by setting $K' = K(\xi)$ through the given K-isomorphism. If we now set

$R = \mathfrak{U}^*$, $\mathfrak{o} = \mathfrak{B}$, $L = K'$ in the Theorem 7 of the Appendix, it is easily seen that all the conditions in the hypothesis are fulfilled; we can therefore conclude that the subring $K'\mathfrak{B}^*$ in $\mathfrak{U}^*$ is an $(K', \mathfrak{B}^*)$-isomorphic image of $\mathfrak{W}^* = K' \times \mathfrak{B}^*$. This shows in particular that $K'\mathfrak{B}^*$ is a complete local ring and its maximal prime ideal is the ideal $K'\mathfrak{v}^*$, which is contained in the ideal $\mathfrak{u}^*$; it follows then that the subring $K'\mathfrak{B}^*$ is equal to its own adherence in $\mathfrak{U}^*$. Therefore, to complete our proof, we only need to show that $\mathfrak{U}^* = K'\mathfrak{B}^*$; for then the two rings $\mathfrak{U}^*$ and $\mathfrak{W}^*$ will be K'-isomorphic and *a fortiori* also K-isomorphic. To prove this equation $\mathfrak{U}^* = K'\mathfrak{B}^*$, it is sufficient to show that the elements of $K'\mathfrak{B}$ are everywhere dense in the local ring $\mathfrak{U}^*$. This means that given any element α of $\mathfrak{U}^*$ and a positive integer s, there is an element β_s of $K'\mathfrak{B}$ such that $\alpha \equiv \beta_s \pmod{\mathfrak{u}^{*s}}$. Now, according to Lemma II, the ideal $\mathfrak{u}$ of $\mathfrak{U}$ has a basis Φ_j, $j = 1, \cdots, h$, which are elements of $\mathfrak{B}$, hence also of $K'\mathfrak{B}$. It is well known that the elements Φ_j, $j = 1, \cdots, h$, constitute then also a basis of the ideal $\mathfrak{u}^*$ in $\mathfrak{U}^*$. Since the elements of K' constitute a complete set of representatives of the residue field $\mathfrak{U}^*/\mathfrak{u}^*$, there exists an element α_1 in K' such that $\alpha \equiv \alpha_1 \pmod{\mathfrak{u}^*}$. We set $\beta_1 = \alpha_1$. Assuming that the element β_{s-1} has already been constructed, we proceed to construct the element β_s. Since the element $\alpha - \beta_{s-1}$ is an element in the ideal $\mathfrak{u}^{*s-1}$, we can express it as a form of degree $s - 1$ in the basis elements Φ_j, $j = 1, \cdots, h$, with coefficients in $\mathfrak{U}^*$. Now, by the case $s = 1$, each of these coefficients is congruent to an element of K' modulo $\mathfrak{u}^*$. If we now substitute every coefficient of the form by the corresponding element in K', we obtain an element α_s in $K'\mathfrak{B}$ such that $\alpha - \beta_{s-1} \equiv \alpha_s \pmod{\mathfrak{u}^{*s}}$. The element $\beta_s = \alpha_s + \beta_{s-1}$ is in the ring $K'\mathfrak{B}$ and satisfies the condition $\alpha \equiv \beta_s \pmod{\mathfrak{u}^{*s}}$. Thus the proof for the case $d = 0$ is complete.

(b) The case $d > 0$. As we have mentioned at the end of Section 2, we shall assume that the affine coordinate system of S_m has been so chosen that the $(m - d)$-dimensional linear variety defined by the equations $X_j = v_j$, $j = 1, \cdots, d$, where the $v_1, \cdots, v_d$ are independent variables over K, intersects the d-cycle $G(\eta)$ properly and finitely. Then, if G_1/K is any K-component of degree n_1 of $G(\eta)$, a generic point (ζ) of G_1/K can be chosen such that the first d coordinates $\zeta_1, \cdots, \zeta_d$ are any given set of d independent variables over K, and the field $K(\zeta_1, \cdots, \zeta_m)$ is an algebraic extension of degree n_1 over the field $K(\zeta_1, \cdots, \zeta_d)$. Since the first d coordinates $x_1, \cdots, x_d$ of the generic point (x) of U/K are in this case independent variables over K, we can therefore set $\zeta_j = x_j$ for $j = 1, \cdots, d$. In other words, each K-component of $G(\eta)$ has a generic point with the first d coordinates $x_1, \cdots, x_d$. This holds in particular for the K-component G'/K

of $G(\eta)$, and we have a generic point (ξ) with $\xi_j = x_j$ for $j = 1, \cdots, d$. Since the variety G'/K is separably generated, we may without loss of generality assume that the affine coordinate system of S_m has been so chosen that $K(\xi)$ is separably algebraic of degree n' over field $K(x_1, \cdots, x_d)$.

Let $\tilde{K} = K(x_1, \cdots, x_d)$, and let $U/\tilde{K}$ and $V/\tilde{K}$ be the extensions of the varieties U/K and V/K respectively over $\tilde{K}$. Let $L/\tilde{K}$ be the linear variety of dimension $m - d$ defined by the d linear equations $X_j = x_j$, $j = 1, \cdots, d$. Then the intersection of $U/\tilde{K}$ with $L/\tilde{K}$ is a variety $\tilde{U}/\tilde{K}$ over $\tilde{K}$, and the point (x) is also a generic point of $\tilde{U}/\tilde{K}$. The point (ξ) is evidently also a point of $\tilde{U}/\tilde{K}$, and it is well known[20] that the quotient ring $Q(\tilde{U}/\tilde{K}, (\xi))$ coincides with the quotient ring $\mathfrak{U} = Q(U/K, (\xi))$. Since the point (x) is a generic point of $\tilde{U}/\tilde{K}$ and the point (y) is also a point of $V/\tilde{K}$, the rational transformation $(y) = T(x)$ of U/K onto V/K can also be regarded as a rational transformation $\tilde{T}$ of the variety $\tilde{U}/\tilde{K}$ into the variety $V/\tilde{K}$. In fact, this rational transformation T is also a transformation of $\tilde{U}/\tilde{K}$ *onto* $V/\tilde{K}$; in other words, the point (y) is not only a generic point of V/K, but also a generic point of $V/\tilde{K}$. To prove this, we only need to show that every generic point (y') of V/K is a specialization of the point (y) over $\tilde{K}$. In fact, since $G(y')$ is a generic d-cycle of the involutionary system $| G(y) |$, it has at least a finite intersection point $(x') = (x_1, \cdots, x_d, x'_{d+1}, \cdots, x'_m)$ with the linear variety $L/\tilde{K}$. The pair of points (x'), (y') is then a specialization of the pair (x), (y) over K; hence, the pair $(x_1, \cdots, x_d)$, (y') is also a specialization of the pair $(x_1, \cdots, x_d)$, (y) over K, and this means that the point (y') is a specialization of (y) over $\tilde{K}$.

It is evident from the definition of the transformation $\tilde{T}$ that to the point (y) of $V/\tilde{K}$ correspond in the inverse transformation $\tilde{T}^{-1}$ exactly the points of the intersection of $L/\tilde{K}$ with the d-cycle $G(y)$. Since each K-component G_1/K of $G(\eta)$ has a generic point of the type $(\zeta) = (x_1, \cdots, x_d, \zeta_{d+1}, \cdots, \zeta_m)$ such that the field $K(\zeta)$ is an algebraic extension of degree n_1 over $\tilde{K}$, it follows that the intersection of G_1/K (considered as a d-cycle) with $L/\tilde{K}$ is a finite 0-cycle $\tilde{G}_1$ of degree n_1, prime rational over $\tilde{K}$; thus we have a 0-dimensional variety $\tilde{G}_1/\tilde{K}$, and the point (ζ) is also a generic point of $\tilde{G}_1/\tilde{K}$. This holds in particular for the intersection $\tilde{G}'/\tilde{K}$ of $G'/\tilde{K}$ with $L/\tilde{K}$; and since the field $\tilde{K}(\xi) = K(\xi)$ is separable over $\tilde{K}$, this 0-dimensional variety $\tilde{G}'/\tilde{K}$ is also separably generated over $\tilde{K}$. Summing over all the K-components of the cycle $G(\eta)$, we conclude that the intersection of $G(\eta)$ with $L/\tilde{K}$ is a finite 0-cycle $\tilde{G}(\eta)$ of degree n, rational over $\tilde{K}$, which contains the variety $\tilde{G}'/\bar{K}$ as a simple $\bar{K}$-component. It is well known that in such a case the intersection of the generic cycle $G(y)$ with $L/\bar{K}$ is also a finite 0-cycle $\tilde{G}(y)$ of degree n, prime rational over $\bar{K}(y)$, and the cycle $\tilde{G}(\eta)$ is the unique specialization of the cycle $\tilde{G}(y)$ over the specialization $(y) \rightarrow (\eta)$ over $\bar{K}$. Thus the rational transformation $\check{T}$ induces in $\bar{U}/\bar{K}$ an involutional system of 0-cycles $|\tilde{G}(y)|$, and the inverse transformation $\check{T}^{-1}$ is defined at the point (η). Since the transformation $\check{T}$ is evidently regular at the point (ξ), it follows then from Lemma I that the quotient ring $\mathfrak{U} = Q(\bar{U}/\bar{K}, (\xi))$ contains the quotient ring $\check{\mathfrak{V}} = Q(V/\bar{K}, (\eta))$ as a sub-space.

---

[20] See Zariski [10], p. 8, Lemma 1.

The variety $V/\bar{K}$ is however not the associated variety of the involutional system $|\tilde{G}(y)|$. To find the associated variety of this system, we only need to observe that the associated form of the generic cycle $\tilde{G}(y)$ is the form $\tilde{F}(Z) = F_{(y)}(Z, u^{(1)}, \cdots, u^{(d)})$, where the form $F_{(y)}(Z, Z^{(1)}, \cdots, Z^{(d)})$ is the associated form of the d-cycle $G(y)$ and the $(u^{(i)})$, $i = 1, \cdots, d$, are defined as follows: For each $i = 1, \cdots, d$, we set $u^{(i)}{}_j = -x_i$, 1, or 0, according as $j = 0$, $j = i$, or $j \neq 0, i$, respectively. As we have mentioned at the end of Section **2**, the form $F_{(y)}(Z, Z^{(1)}, \cdots, Z^{(d)})$ is assumed to be so normalized that the coefficient of the term ω_0 is 1. It follows then the coefficient of the term $Z_0{}^n$ in the form $\tilde{F}(Z)$ is also equal to 1, so that $\tilde{F}(Z)$ is already in the normalized form. Let $(\tilde{y})$ be the set of coefficients (except the coefficient 1 of the term $Z_0{}^n$) of this form $\tilde{F}(Z)$, ordered in some arbitrary but fixed way, then the variety $\tilde{V}/\bar{K}$ determined by the point $(\tilde{y})$ over $\bar{K}$ is the associated variety of the involutional system $|\tilde{G}(y)|$; and we can now write $\tilde{G}(\tilde{y})$ instead of $\tilde{G}(y)$ to indicate its dependence on the point $(\tilde{y})$. It is clear that the coordinates of $(\tilde{y})$ are linear combination of the coordinates of (y); this means that there is a projection P of $V/\bar{K}$ onto $\tilde{V}/\bar{K}$ and that this projection P is regular at every finite point of $V/\bar{K}$, hence in particular at the point (η). We shall denote by $(\bar{\eta})$ the corresponding point of (η) under this projection P, and write $\tilde{G}(\bar{\eta})$ instead of $\tilde{G}(\eta)$ to indicate its dependence on the point $(\bar{\eta})$ of $\tilde{V}/\bar{K}$. Since the hypothesis of Lemma I is evidently satisfied by P at the point (η), it follows that the quotient ring $\check{\mathfrak{V}} = Q(V/\bar{K}, (\eta))$ contains the quotient ring $\tilde{\mathfrak{V}} = Q(\tilde{V}/\bar{K}, (\bar{\eta}))$ of $\tilde{V}/\bar{K}$ at the point $(\bar{\eta})$ as a sub-space; and this means that $\tilde{\mathfrak{V}}$ is also a sub-space of the local ring $\mathfrak{U}$.

Let $\tilde{T}$ be the associated rational transformation of the involutional system $|\tilde{G}(\tilde{y})|$; it is easily seen that $\tilde{T}$ is the product of the rational transformation $\check{T}$ of $\bar{U}/\bar{K}$ onto $V/\bar{K}$ and the projection P of $V/\bar{K}$ onto $\tilde{V}/\bar{K}$. Since the transformation $\check{T}$ is regular at the point (ξ) and the image point on $V/\bar{K}$ is the point (η), and since the projection P is regular at the point (η) and the image point on $\tilde{V}/\tilde{K}$ is the point $(\tilde{\eta})$, it follows that the associated transformation $\tilde{T}$ is regular at the point (ξ) and the image point on $\tilde{V}/\tilde{K}$ is the point $(\tilde{\eta})$. Moreover, the 0-dimensional variety $\tilde{G}'/\tilde{K}$ is a simple $\tilde{K}$-component of the cycle $\tilde{G}(\tilde{\eta})$ and is separably generated; hence the hypothesis of Lemma II is fulfilled by the involutional system $|\tilde{G}(\bar{y})|$ on the variety $\bar{U}/\tilde{K}$ with the associated variety $\tilde{V}/\tilde{K}$ and the associated transformation $\tilde{T}$. It follows then that the maximal prime ideal $\mathfrak{u}$ in the quotient ring $\mathfrak{U} = Q(\bar{U}/\tilde{K}, (\xi))$ has a basis consisting of elements in the subring $\tilde{\mathfrak{V}}$ and hence also in the subring $\mathfrak{V}$ of $\mathfrak{U}$.

Finally, let $\mathfrak{W} = Q(W/K, (\eta, \xi))$ be the quotient ring of W/K at the point (η, ξ). It is clear that $\mathfrak{W}$ is also the quotient ring $Q(V/K(\xi), (\eta))$ of the extension variety $V/K(\xi)$ of V/K over $K(\xi)$ at the point (η), and the variety $V/K(\xi)$ is also the extension variety of $V/\tilde{K}$ over $K(\xi)$; this means that the local ring $\mathfrak{W}$ is the extension of the local ring $\tilde{\mathfrak{V}}$ over the field $K(\xi)$. Since $K(\xi)$ is a finite extension of $\tilde{K}$ and since $\tilde{V}$ contains $\tilde{K}$ as a coefficient field, it follows that $\mathfrak{W}^* = K(\xi) \times \tilde{\mathfrak{V}}^*$. Furthermore, since the field $K(\xi)$ is analytically disjoint with respect to the quotient ring $\mathfrak{V} = Q(V/K, (\eta))$ over the field K, and since $\tilde{\mathfrak{V}}$ is the extension of $\mathfrak{V}$ over the subfield $\tilde{K}$ of $K(\xi)$, it follows (Appendix, Theorem 6) that $K(\xi)$ is also analytically disjoint with respect to $\tilde{\mathfrak{V}}$ over the field $\tilde{K}$.

We are now ready to apply the method of (a). Consider the completion $\mathfrak{U}^*$ of $\mathfrak{U}$; it contains the completion $\tilde{\mathfrak{V}}^*$ of $\tilde{\mathfrak{V}}$ as a subring. Since $K(\xi)$ is separable over $\tilde{K}$, the local ring $\mathfrak{U}$ contains $\tilde{K}$ as a basic field; hence the completion $\mathfrak{U}^*$ contains a coefficient field $\tilde{K}'$ over $\tilde{K}$, which is $\tilde{K}$-isomorphic to the field $K(\xi)$. Here again the field $\tilde{K}'$ is an "abstract" field, but we can just as before identify it with $K(\xi)$ by setting $\tilde{K}' = K(\xi)$ through the given $\tilde{K}$-isomorphism. If we set $R = \mathfrak{U}^*$, $\mathfrak{o} = \tilde{\mathfrak{V}}$, $L = \tilde{K}'$, $K = \tilde{K}$ in Theorem 7 of the Appendix, then all the condition in the hypothesis are fulfilled; we can therefore conclude that the subring $\tilde{K}'\tilde{\mathfrak{V}}^*$ in $\mathfrak{U}^*$ is a $(\tilde{K}', \tilde{\mathfrak{V}}^*)$-isomorphic image of $\mathfrak{W}^* = \tilde{K}' \times \tilde{\mathfrak{V}}^*$. Observing that the maximal prime ideal in $\mathfrak{U}^*$ has a basis consisting of elements in $\tilde{\mathfrak{V}}$, we can repeat the same argument as in the last part of (a) and conclude that $\mathfrak{U}^* = \tilde{K}'\tilde{\mathfrak{V}}^*$. Therefore the two rings $\mathfrak{U}^*$ and $\mathfrak{W}^*$ are $\tilde{K}'$-isomorphic and *a fortiori* also K-isomorphic with each other. This completes the proof of our theorem.

APPENDIX.*

On the Extensions of Local Domains.

In this appendix we shall develop some notions and results concerning the extension of a local domain over a field, which are used in the text, but are not directly related to the problem considered there, so that we consider it more appropriate to present them separately. Theorems 1-6 are not essentially new; in fact, most of them are contained implicitly in the work of Chevalley, to which we shall refer for some of the proofs. However, for our present purpose, we have to develop the subject in a somewhat different way. The one new result is Theorem 7, which plays an essential part in the proof of the main theorem in the text. As a matter of terminology, we stress here that the expression "field" will be used in this appendix in its usual general sense as meaning an abstract field, not necessarily a subfield of a certain "universal domain."

Let R_1 and R_2 be two (commutative) rings which contain a field K as common subring. We shall consider the Kronecker product $R_1 \times R_2$ of R_1 and R_2 over the field K, which is a ring containing both R_1 and R_2 as subrings. The Kronecker product $R_1 \times R_2$ has the following characteristic property: Let R' be any ring containing K, and let R'_1 and R'_2 be two subrings of R', both containing K; then if R'_1 and R'_2 are K-isomorphic to R_1 and R_2 respectively, then the subring[21] $R'_1R'_2$ in R' is a K-homomorphic image of $R_1 \times R_2$. Furthermore, this homomorphism will be an isomorphism if and only if R'_1 and R'_2 are linearly disjoint over K. We recall here that R'_1 and R'_2 are linearly disjoint over K, if every set of linearly independent elements of R'_1 over K is still such over R'_2; and that when this is so, then every set of linearly independent elements of R'_2 over K is also still such over R'_1. If we identify the rings R'_1 and R'_2 with the rings R_1 and R_2 respectively, as we shall often do, then the homomorphism of $R_1 \times R_2$ onto the subring R_1R_2 in R' is not only a K-homomorphism, but also an (R_1, R_2)-homomorphism in the sense that both R_1 and R_2 are left invariant by the homomorphism.

Let L and M be two fields which contain a common subfield K. We shall say that L and M are *algebraically disjoint* over K, if the prime ideal in the polynomial ring $K[X_1, \cdots, X_n]$ determined by every set of n elements in L generates a prime ideal in the polynomial ring $M[X_1, \cdots, X_n]$. It is easily seen that in case both L and M are subfields of a larger field such that L and M are independent over K, then they are algebraically disjoint over K if and only if they are linearly disjoint over K. On the other hand, given any two fields L and M having exactly a common subfield K, there always exists a field N which contains L and a subfield M', K-isomorphic to M, such that L and M' are independent over K. If L and M are algebraically disjoint over K, then L and M' are also algebraically disjoint over K and hence also linearly disjoint over K; and since in this case the intersection of L and M' is exactly the field K, we can identify M' with M, so that N contains both L and M as subfields. Thus there is no essential difference between algebraical disjointness and linear disjointness; the former is a natural generalization of the latter to two arbitrary fields over K and hence more suitable to use in those investigations in which it is not possible or convenient to impose a fixed "universal domain" in advance.

---

* Added January 20, 1950.

[21] We shall denote by $R'_1R'_2$ the subring of R' consisting of all the sums of products of elements in R'_1 and R'_2. Thus, in case of two subfields L and M of a larger field, the ring LM is in general *not* the compositum of L and M.

The above described relation between algebraical disjointness and the linear disjointness enables us to carry over most of the results about the linear disjointness to algebraical disjointness. Thus, in particular, we can conclude that algebraical disjointness is a symmetrical relation between the two fields. Furthermore, we can define a field L as a regular extension of K, if L and $\bar{K}$ (the algebraical closure of K) are algebraically disjoint over K. It follows then from a well known result[22] that if L is a regular extension of K and M is any field containing K, then L and M are algebraically disjoint over K.

The definition of algebraical disjointness can be extended to two integral domains. We shall say that two integral domains R_1 and R_2, both containing the field K as subring, are algebraically disjoint over K, if their corresponding fields of quotients L and M are algebraically disjoint over K.

THEOREM 1. *Let R_1 and R_2 be two integral domains which both contain the field K as subring. Then the Kronecker product $R_1 \times R_2$ of R_1 and R_2 over K is an integral domain if and only if R_1 and R_2 are algebraically disjoint over K.*

Proof. If $R_1 \times R_2$ is an integral domain, then the fields L and M will be subfields in the field of quotients of $R_1 \times R_2$, and since R_1 and R_2 are linearly disjoint over K, it follows that L and M are also linearly disjoint over K and hence algebraically disjoint over K. Conversely, if L and M are algebraically disjoint over K, let N be a field containing both L and M as subfields, such that L and M are linearly disjoint over K. Then the subring LM in N is a homomorphic image of $L \times M$; and since L and M are linearly disjoint over K, this homomorphism is an isomorphism. Since LM is an integral domain, it follows that $L \times M$ is an integral domain, and hence $R_1 \times R_2$ is also an integral domain.

---

[22] See Weil [8], p. 18, Theorem 5.

THEOREM 2. *Let R be an integral domain containing the field K as subring, and let L be an algebraic extension of K. If L and R are algebraically disjoint over K, then any R-homomorphism of $L \times R$ is an isomorphism.*

Proof. In fact, if M is the field of quotients of R, then any R-homomorphism of $L \times R$ can be extended to an M-homomorphism of the ring of quotients of $L \times R$ with respect to the multiplicatively closed system R, which is the ring $L \times M$. Since L is algebraic, $L \times M$ is a field, and hence the homomorphism must be an isomorphism.

Following Cohen,[23] we shall define a generalized local ring $\mathfrak{o}$ as a (commutative) ring with an element 1 in which (1) the set $\mathfrak{m}$ of all non-units is an ideal with finite basis, and (2) $\bigcap_{s=1}^{\infty} \mathfrak{m}^s = (0)$. If a generalized local ring is a Noetherian ring, then it is a local ring in the sense of Krull. It is known [24] that a complete generalized local ring is a local ring, so that in any case the completion $\mathfrak{o}^*$ of $\mathfrak{o}$ is a local ring. We shall speak of a local domain or generalized local domain if the ring has no zero-divisors.

If a generalized local ring $\mathfrak{o}$ contains a field K as subring, then no element of K is contained in the maximal prime ideal $\mathfrak{m}$ of $\mathfrak{o}$, and hence the residue ring $\mathfrak{o}/\mathfrak{m}$ can be considered as an extension field of K. Furthermore, the completion $\mathfrak{o}^*$ of $\mathfrak{o}$ also contains the field K as subring, and the residue ring $\mathfrak{o}^*/\mathfrak{o}^*\mathfrak{m}$, which is K-isomorphic to $\mathfrak{o}/\mathfrak{m}$, can also be considered as an extension field of K. Any field K' in the ring $\mathfrak{o}$ which is a complete system of representatives of the residue ring $\mathfrak{o}/\mathfrak{m}$ is called a *coefficient field* in $\mathfrak{o}$; and if the field K' contains the field K as a subfield, then K' is said to be a *coefficient field over* K in $\mathfrak{o}$. It has been shown by Cohen [25] that while the ring $\mathfrak{o}$ in general does not contain a coefficient field, the completion $\mathfrak{o}^*$ of $\mathfrak{o}$ always contains a coefficient field, provided $\mathfrak{o}/\mathfrak{m}$ has the same characteristic as $\mathfrak{o}$; however, even in $\mathfrak{o}^*$ there need not always exist a coefficient field over any given field K. We shall say that a generalized local ring $\mathfrak{o}$ contains a *basic field* [26] K if $\mathfrak{o}^*$ contains a coefficient field over K.

---

[23] Cohen [6].

[24] See Cohen [6], p. 61, Theorem 3.

[25] See Cohen [6], p. 72, Theorem 9.

[26] It is to be noticed that we use the expression "basic field" in a quite different sense from that of Chevalley in [2].

THEOREM 3. *Let U/K be a variety in S_m, and let (ξ) be a point in U/k such that $K(\xi)$ is a separably generated extension of K. Then the quotient ring $Q(U/K, (\xi))$ contains K as a basic field.*

Proof. Let r be the degree of transcendency of $K(\xi)$ over K. It is well known that we can always find a suitable affine transformation in S_m, such that after such a transformation the first r coördinates $\xi_1, \cdots, \xi_r$ of (ξ) are independent variables over K, and the field $K(\xi)$ is separably algebraic over $K(\xi_1, \cdots, \xi_r)$. This implies that if $(x) = (x_1, \cdots, x_m)$ is a generic point of U/K, then the $x_1, \cdots, x_r$ are independent variables over K and the quotient ring $Q(U/K, (\xi))$ contains the field $K(x_1, \cdots, x_r)$. The residue ring of $Q(U/K, (\xi))$ over its maximal prime ideal, being isomorphic to the field $K(\xi)$, is then a separably algebraic extension of $K(x_1, \cdots, x_r)$; and it is well known[27] that in this case the completion of $Q(U/K, (\xi))$ contains a coefficient field over $K(x_1, \cdots, x_r)$, which is then evidently also a coefficient field over K.

Let $\mathfrak{o}$ be a generalized local domain with a basic field K, $\mathfrak{m}$ be the maximal prime ideal in $\mathfrak{o}$, and M be the quotient field of $\mathfrak{o}$. Let L be any field containing K, such that both L and M as well as L and $\mathfrak{o}/\mathfrak{m}$ are algebraically disjoint over K. According to Theorem 1, the Kronecker product $L \times \mathfrak{o}$ of L and $\mathfrak{o}$ over K is an integral domain, and it contains $\mathfrak{o}$ and L as two linearly disjoint subrings over K. It can be easily shown that $(L \times \mathfrak{o})\mathfrak{q} \cap \mathfrak{o} = \mathfrak{q}$ for every ideal $\mathfrak{q}$ in $\mathfrak{o}$, and that $\bigcap_i (L \times \mathfrak{o})\mathfrak{q}_i = (L \times \mathfrak{o})(\bigcap_i \mathfrak{q}_i)$ for every set of ideals $\mathfrak{q}_i$ in $\mathfrak{o}$. In particular, we have the relations $\bigcap_{s=1}^{\infty} (L \times \mathfrak{o})\mathfrak{m}^s = (L \times \mathfrak{o})(\bigcap_{s=1}^{\infty} \mathfrak{m}^s) = (0)$ and $(L \times \mathfrak{o})\mathfrak{m}^s \cap \mathfrak{o} = \mathfrak{m}^s$; which shows that if we consider $L \times \mathfrak{o}$ as a (generalized) $(L \times \mathfrak{o})\mathfrak{m}$-adic ring in the sense of Zariski,[28] then $L \times \mathfrak{o}$ contains $\mathfrak{o}$ as a subspace. Moreover, it is easily seen that the residue ring $L \times \mathfrak{o}/(L \times \mathfrak{o})\mathfrak{m}$ is isomorphic to the ring $L \times (\mathfrak{o}/\mathfrak{m})$ which is according to Theorem 1 an integral domain; this shows that $(L \times \mathfrak{o})\mathfrak{m}$ is a prime ideal in $L \times \mathfrak{o}$. Let $\mathfrak{o}_L$ be the quotient ring of $L \times \mathfrak{o}$ with respect to the multiplicatively closed system $L \times \mathfrak{o} - (L \times \mathfrak{o})\mathfrak{m}$; then $\mathfrak{o}_L\mathfrak{m}$ is the ideal of non-units in $\mathfrak{o}_L$ and has obviously a finite basis. We shall show now that $\mathfrak{o}_L$ is a generalized local domain and contains $\mathfrak{o}$ as a subspace; to do this it is sufficient to show that $\bigcap_{s=1}^{\infty} \mathfrak{o}_L\mathfrak{m}^s = (0)$ and $\mathfrak{o}_L\mathfrak{m}^s \cap \mathfrak{o} = \mathfrak{m}^s$.

---

[27] See Chevalley [2], p. 701, Proposition 3.

[28] See Zariski [9]; we add the word "generalized" to indicate the fact that the ring is not necessarily Noetherian.

In case K is a coefficient field of $\mathfrak{o}$, the residue ring $L \times \mathfrak{o}/(L \times \mathfrak{o})\mathfrak{m}$ is isomorphic to the field L; this implies that the prime ideal $(L \times \mathfrak{o})\mathfrak{m}$ is maximal in $L \times \mathfrak{o}$ and hence the ideals $(L \times \mathfrak{o})\mathfrak{m}^s$ are all primary. It follows then that $\bigcap_{s=1}^{\infty} \mathfrak{o}_L\mathfrak{m}^s = \mathfrak{o}_L(\bigcap_{s=1}^{\infty} (L \times \mathfrak{o})\mathfrak{m}^s) = \mathfrak{o}_L(\bigcap_{s=1}^{\infty} \mathfrak{m}^s) = (0)$ and $\mathfrak{o}_L\mathfrak{m}^s \cap \mathfrak{o} = \mathfrak{o}_L\mathfrak{m}^s \cap L \times \mathfrak{o} \cap \mathfrak{o} = (L \times \mathfrak{o})\mathfrak{m}^s \cap \mathfrak{o} = \mathfrak{m}^s$. If K is not a coefficient field of $\mathfrak{o}$, then the completion $\mathfrak{o}^*$ of $\mathfrak{o}$ contains a coefficient K' over K. Since K' is K-isomorphic to $\mathfrak{o}/\mathfrak{m}$, the fields L and K' are algebraically disjoint over K. Therefore the Kronecker product $L \times K'$ of L and K' over K is an integral domain; let L' be the field of quotients of $L \times K'$. Since K' is a coefficient field of $\mathfrak{o}^*$ and L' contains K', we can apply to L' and $\mathfrak{o}^*$ over the field K' the case we have just proved and conclude that $\bigcap_{s=1}^{\infty} (\mathfrak{o}^*)_{L'}\mathfrak{m}^s = (0)$ and $(\mathfrak{o}^*)_{L'}\mathfrak{m}^s \cap \mathfrak{o}^* = \mathfrak{o}^*\mathfrak{m}^s$. Since the ring $(\mathfrak{o}^*)_{L'}$ contain $\mathfrak{o}_L$ as a subring, it follows then that $\bigcap_{s=1}^{\infty} \mathfrak{o}_L\mathfrak{m}^s \subset \mathfrak{o}_L \cap (\bigcap_{s=1}^{\infty} (\mathfrak{o}^*)_{L'}\mathfrak{m}^s) = (0)$ and $\mathfrak{o}_L\mathfrak{m}^s \cap \mathfrak{o} \subset (\mathfrak{o}^*)_{L'}\mathfrak{m}^s \cap \mathfrak{o} = \mathfrak{o}^*\mathfrak{m}^s \cap \mathfrak{o} = \mathfrak{m}^s$, and hence $\bigcap_{s=1}^{\infty} \mathfrak{o}_L\mathfrak{m}^s = (0)$ and $\mathfrak{o}_L\mathfrak{m}^s \cap \mathfrak{o} = \mathfrak{m}^s$.

Thus we have shown that in any case the ring $\mathfrak{o}_L$ is a generalized local domain and contains the generalized local domain $\mathfrak{o}$ as a subspace. We shall call $\mathfrak{o}_L$ the *extension* of $\mathfrak{o}$ over the field L. We remark further that in case L is a finitely generated extension of K (which is the only case of interest for our present purpose), it can be shown that $\mathfrak{o}_L$ is a Noetherian ring if $\mathfrak{o}$ is such; in other words, in this case $\mathfrak{o}_L$ is a local domain if $\mathfrak{o}$ is a local domain. However, this additional restriction is not necessary for our considerations here. It is easily seen that if a variety U/K has an extension over a field L containing K and if (ξ) is a point of U/K such that $K(\xi)$ and L are linearly disjoint over K, then the quotient ring $Q(U/L, (\xi))$ is the extension of the quotient ring $Q(U/K, (\xi))$ over the field L.

Consider the completion $(\mathfrak{o}_L)^*$ of the generalized local domain $\mathfrak{o}_L$; $(\mathfrak{o}_L)^*$ is a local ring and contains the local ring $\mathfrak{o}^*$ as a subring. Let $\mathfrak{p}_1, \cdots, \mathfrak{p}_n$ be the minimal (unimbedded) prime divisors of the zero ideal in $\mathfrak{o}^*$. Then for each $i = 1, \cdots, n$, the residue ring $\mathfrak{o}^*/\mathfrak{p}_i$ is a complete local domain, and the residue ring $(\mathfrak{o}_L)^*/(\mathfrak{o}_L)^*\mathfrak{p}_i$ is a complete local ring and contains $\mathfrak{o}^*/\mathfrak{p}_i$ as a subring. We shall say that the field L is *analytically disjoint* with respect to the generalized local domain $\mathfrak{o}$ over the field K, if the ideals $(\mathfrak{o}_L)^*\mathfrak{p}_i$, $i = 1, \cdots, n$, are all prime ideals in $(\mathfrak{o}_L)^*$, so that the rings $(\mathfrak{o}_L)^*/(\mathfrak{o}_L)^*\mathfrak{p}_i$, $i = 1, \cdots, n$, are all complete local domains.

Let $\mathfrak{q}$ be any ideal in $\mathfrak{o}^*$; we are concerned with the relations between the complete local ring $(\mathfrak{o}_L)^*/(\mathfrak{o}_L)^*\mathfrak{q}$ and the complete local ring $\mathfrak{o}^*/\mathfrak{q}$. Let K' be a coefficient field over K in the complete local ring $\mathfrak{o}^*$, and let L' be the field of quotients of the Kronecker product $L \times K'$ over K. Then the Kronecker product $L' \times \mathfrak{o}^*$ of L' and $\mathfrak{o}^*$ over K' is an $\mathfrak{m}$-adic ring, and the residue ring $L' \times \mathfrak{o}^*/(L' \times \mathfrak{o}^*)\mathfrak{m}$ is isomorphic with the subfield L' of $L' \times \mathfrak{o}^*$. It is then easily seen that every element in the set $L' \times \mathfrak{o}^* - (L' \times \mathfrak{o}^*)\mathfrak{m}$ is $\equiv 1$ (mod $(L' \times \mathfrak{o}^*)\mathfrak{m}$), from which it follows[29] that the $(L' \times \mathfrak{o}^*)\mathfrak{m}$-adic ring $L' \times \mathfrak{o}^*$ and the local ring $(\mathfrak{o}^*)_{L'}$ have the same completion. Since the completion of $(\mathfrak{o}^*)_{L'}$ is evidently equal to $(\mathfrak{o}_L)^*$, we have therefore $(\mathfrak{o}_L)^* = (L' \times \mathfrak{o}^*)^*$. If now $\mathfrak{q}$ is any ideal $(\neq \mathfrak{o}^*)$ in $\mathfrak{o}^*$, then the residue ring $L' \times \mathfrak{o}^*/(L' \times \mathfrak{o}^*)\mathfrak{q}$ contains L' and $\mathfrak{o}^*/\mathfrak{q}$ as subrings and is $(L', \mathfrak{o}^*/\mathfrak{q})$-isomorphic with the Kronecker product $L' \times (\mathfrak{o}^*/\mathfrak{q})$ over K'. Since the completion of $L' \times \mathfrak{o}^*/(L' \times \mathfrak{o}^*)\mathfrak{q}$ is the residue ring $(L' \times \mathfrak{o}^*)^*/(L' \times \mathfrak{o}^*)^*\mathfrak{q} = (\mathfrak{o}_L)^*/(\mathfrak{o}_L)^*\mathfrak{q}$, we conclude therefore that the two complete local rings $(\mathfrak{o}_L)^*/(\mathfrak{o}_L)^*\mathfrak{q}$ and $(L' \times (\mathfrak{o}^*/\mathfrak{q}))^*$ are $(L', \mathfrak{o}^*/\mathfrak{q})$-isomorphic with each other. In other words, the complete local ring $(\mathfrak{o}_L)^*/(\mathfrak{o}_L)^*\mathfrak{q}$ contains the Kronecker product $L' \times (\mathfrak{o}^*/\mathfrak{q})$ of the two subrings L' and $\mathfrak{o}^*/\mathfrak{q}$ over K' as a subspace and coincides with the adherence of the latter.

The following two theorems (in somewhat different form) are due to Chevalley:

THEOREM 4. *If* $\mathfrak{o}$ *is a generalized local domain with a basic field* K *and if the residue field of* $\mathfrak{o}$ *is a finitely generated extension of* K, *then any regular extension of* K *is analytically disjoint with respect to* $\mathfrak{o}$ *over* K.

THEOREM 5. *If* $\mathfrak{o}$ *is a generalized local domain with a coefficient field* K, *and if* $\bar{K}$ *is analytically disjoint with respect to* $\mathfrak{o}$ *over* K, *then any field containing* K *is analytically disjoint with respect to* $\mathfrak{o}$ *over* K.

For proof we shall refer to Chevalley [2], § 4. It is sufficient to observe that if we set $(L' \times (\mathfrak{o}^*/\mathfrak{p}_i))^*$, $\mathfrak{o}^*/\mathfrak{p}_i$, L', K' equal to $\mathfrak{O}$, $\mathfrak{o}$, Z, K in the notation of Chevalley [2], p. 77, then all the conditions there are fulfilled. If L is a regular extension of K (i.e. K is "strongly algebraically closed" in L, in the terminology of Chevalley), then L' is also a regular extension of K', and Theorem 4 then follows from [2], p. 78, Proposition 9a and the remark immediately after it. Theorem 5 follows from [2], p. 80, Proposition 10b.

We shall also mention the following rather trivial theorem:

---

[29] See Zariski [9], p. 183. That our ring $L' \times \mathfrak{o}^*$ is not necessarily Noetherian is immaterial here.

THEOREM 6. *Let* $\mathfrak{o}$ *be a generalized local domain containing a field* K, *and let* L *and* L' *be two fields such that* $L \supset L' \supset K$. *If* L *is analytically disjoint with respect to* $\mathfrak{o}$ *over* K, *then* $\mathfrak{o}_{L'}$ *is defined and* L *is analytically disjoint with respect to* $\mathfrak{o}_{L'}$ *over* L'.

For, we have $(\mathfrak{o}_L)^* \supset (\mathfrak{o}_{L'})^* \supset \mathfrak{o}^*$ and $(\mathfrak{o}_L)^*\mathfrak{p}_i \cap (\mathfrak{o}_{L'})^* = (\mathfrak{o}_{L'})^*\mathfrak{p}_i$ for every i; hence if the ideals $(\mathfrak{o}_L)^*\mathfrak{p}_i$ are prime in $(\mathfrak{o}_L)^*$, the ideals $(\mathfrak{o}_{L'})^*\mathfrak{p}_i$ must also be prime in $(\mathfrak{o}_{L'})^*$.

Let $\mathfrak{o}$ be a generalized local domain with a coefficient field K, and let L be a field containing K such that the extension $\mathfrak{o}_L$ of $\mathfrak{o}$ over L is defined. The residue of $L \times \mathfrak{o}/(L \times \mathfrak{o})\mathfrak{m}$ is in this case isomorphic to L, and hence every element in the multiplicatively closed system $L \times \mathfrak{o} - (L \times \mathfrak{o})\mathfrak{m}$ is $\equiv 1$ $(\mathrm{mod}(L \times \mathfrak{o})\mathfrak{m})$. It follows then that the $(L \times \mathfrak{o})\mathfrak{m}$-adic ring $L \times \mathfrak{o}$ and the local ring $\mathfrak{o}_L$ have the same completion; that is, we have $(L \times \mathfrak{o})^* = (\mathfrak{o}_L)^*$ or also $(L \times \mathfrak{o}^*)^* = (\mathfrak{o}_L)^*$, since it is evident that $(L \times \mathfrak{o})^* = (L \times \mathfrak{o}^*)^*$. In case L is a finite algebraic extension of K, we can even write $(\mathfrak{o}_L)^* = L \times \mathfrak{o}^*$; for, in this case the $\mathfrak{m}$-adic ring $L \times \mathfrak{o}^*$ is already complete and hence coincides with $(L \times \mathfrak{o}^*)^*$.

THEOREM 7. *Let* $\mathfrak{o}$ *be a generalized local domain with a coefficient field* K, *such that its completion* $\mathfrak{o}^*$ *has no nilpotent elements. Let* R *be a ring containing* $\mathfrak{o}^*$ *as a subring, and let* L *be a field in* R *such that* L *is a finite algebraic extension of* K. *If* L *is analytically disjoint with respect to* $\mathfrak{o}$ *over* K, *then the subring* $L\mathfrak{o}^*$ *in* R *is* $(L, \mathfrak{o}^*)$*-isomorphic to* $(\mathfrak{o}_L)^*$.

Proof. It is clear that the subring $L\mathfrak{o}^*$ in R is an $(L, \mathfrak{o}^*)$-homomorphic image of $(\mathfrak{o}_L)^* = L \times \mathfrak{o}^*$. Therefore, in order to prove our theorem, it is only necessary to show that the subrings L and $\mathfrak{o}^*$ in R are linearly disjoint over K. Let $\mathfrak{p}_1, \cdots, \mathfrak{p}_n$ be the prime divisors of the zero ideal in $\mathfrak{o}^*$; the assumption that L is analytically disjoint with respect to $\mathfrak{o}$ implies that each ring $L \times \mathfrak{o}^*/(L \times \mathfrak{o}^*)\mathfrak{p}_i$ is an integral domain. Since the ring $L \times \mathfrak{o}^*/(L \times \mathfrak{o}^*)\mathfrak{p}_i$ is $(L, \mathfrak{o}^*/\mathfrak{p}_i)$-isomorphic to the ring $L \times (\mathfrak{o}^*/\mathfrak{p}_i)$, therefore $L \times (\mathfrak{o}^*/\mathfrak{p}_i)$ is also an integral domain; it follows then from Theorem 2 that any $(\mathfrak{o}^*/\mathfrak{p}_i)$-homomorphism of $L \times (\mathfrak{o}^*/\mathfrak{p}_i)$ is an isomorphism. Since every element in $\mathfrak{p}_i$ is annulled by every element in $\bigcap_{j \neq i} \mathfrak{p}_j$, it follows that every element in $(L\mathfrak{o}^*)\mathfrak{p}_i \cap \mathfrak{o}^*$ is also annulled by every element in $\bigcap_{j \neq i} \mathfrak{p}_j$; and since $(L\mathfrak{o}^*)\mathfrak{p}_i \cap \mathfrak{o}^* \supset \mathfrak{p}_i$ and since the ring $\mathfrak{o}^*$ has no imbedded prime divisors of the zero ideal, it follows that $(L\mathfrak{o}^*)\mathfrak{p}_i \cap \mathfrak{o}^* = \mathfrak{p}_i$. It follows then that the residue ring $L\mathfrak{o}^*/(L\mathfrak{o}^*)\mathfrak{p}_i$ contains L and $\mathfrak{o}^*/\mathfrak{p}_i$ as subrings, and it is an $(L, \mathfrak{o}^*/\mathfrak{p}_i)$-homomorphic image of $L \times (\mathfrak{o}^*/\mathfrak{p}_i)$ and hence also an $(L, \mathfrak{o}^*/\mathfrak{p}_i)$-

isomorphic image of $L \times (\mathfrak{o}^*/\mathfrak{p}_i)$. Thus each $L\mathfrak{o}^*/(L\mathfrak{o}^*)\mathfrak{p}_i$ is a complete local domain and contains L and $\mathfrak{o}^*/\mathfrak{p}_i$ as two linearly disjoint subrings over K.

Let $u_1, \cdots, u_m$ be a set of linearly independent elements over K in $\mathfrak{o}^*$, and denote by $(u_1, \cdots, u_m)$ the linear space determined by this set over K. If l_1 is the dimension of the sub-space of $(u_1, \cdots, u_m)$ which is contained in the ideal $\mathfrak{p}_1$, then we can replace the set $u_1, \cdots, u_m$ by a linearly equivalent set over K, such that the space $(u_1, \cdots, u_{l_1})$ is in $\mathfrak{p}_1$ and the space $(u_{l_1+1}, \cdots, u_m)$ contains no elements in $\mathfrak{p}_i$. If l_2 is the dimension of the subspace of $(u_1, \cdots, u_{l_1})$ which is contained in $\mathfrak{p}_2$, then we can replace the set $u_1, \cdots, u_{l_1}$ by a linearly equivalent set over K, such that the space $(u_1, \cdots, u_{l_2})$ is in $\mathfrak{p}_1 \cap \mathfrak{p}_2$ and the space $(u_{l_2+1}, \cdots, u_{l_1})$ contains no elements in $\mathfrak{p}_2$. Thus we can proceed until in the final stage we have a set $u_1, \cdots, u_{l_{n-1}}$ such that $(u_1, \cdots, u_{l_{n-1}})$ is contained in $\mathfrak{p}_1 \cap \cdots \cap \mathfrak{p}_{n-1}$; here the process must stop and the space $(u_1, \cdots, u_{l_{n-1}})$ contains no elements in $\mathfrak{p}_n$, for we have $\bigcap_{i=1}^{n} \mathfrak{p}_i = (0)$ on account of the absence of nilpotent elements in $\mathfrak{o}^*$. Adding for the sake of convenience an element $u_0 = 0$, our final results will be a set $u_0, \cdots, u_m$, which is linearly equivalent over K to the original set, but has the property that for each $i = 0, \cdots, n$, the space $(u_0, \cdots, u_{l_i})$ is contained in $\mathfrak{p}_1 \cap \cdots \cap \mathfrak{p}_i$ and the space $(u_{l_i+1}, \cdots, u_{l_{i-1}})$ contains no elements in $\mathfrak{p}_i$ (where we set $l_0 = m$ and $l_n = 0$). We shall now show that if there is a linear relation $\sum_{j=1}^{m} c_j u_j = 0$ with coefficients c_j in L, then all the c_j must vanish. Assume that it has already been proved that $c_j = 0$ for all $j > l_{i-1}$, we shall show that $c_j = 0$ for all $j > l_i$. In fact, if not all $c_{l_i+1}, \cdots, c_{l_{i-1}}$ are zero, then the relation $\sum_{j=1}^{l_{i-1}} c_j u_j = 0$ will induce a linear relation $\sum_{j=l_i+1}^{l_{i-1}} c_j u_j = 0$ in $L\mathfrak{o}^*/(L\mathfrak{o}^*)\mathfrak{p}_i$ between the $u_{l_i+1}, \cdots, u_{l_{i-1}}$, considered as elements in $\mathfrak{o}^*/\mathfrak{p}_i$. Since in $L\mathfrak{o}^*/(L\mathfrak{o}^*)\mathfrak{p}_i$ the subrings L and $\mathfrak{o}^*/\mathfrak{p}_i$ are linearly disjoint over K, it follows that the $u_{l_i+1}, \cdots, u_{l_{i-1}}$ must be linearly dependent over K in $\mathfrak{o}^*/\mathfrak{p}_i$; this means that there is a linear combination of the $u_{l_i+1}, \cdots, u_{l_{i-1}}$ with coefficients in K which is an element in $\mathfrak{p}_i$, in contradiction to the fact that the space $(u_{l_i+1}, \cdots, u_{l_{i-1}})$ contains no elements in $\mathfrak{p}_i$. Thus we have shown by induction that all the $c_1, \cdots, c_m$ must vanish, and hence the elements $u_1, \cdots, u_m$ are linearly independent over L. This concludes the proof of our theorem.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

1. C. Chevalley, "On the theory of local rings," *Annals of Mathematics*, vol. 44 (1943), pp. 690-708.

2. ———, "Some properties of ideals in rings of power series," *Transactions of the Americal Mathematical Society*, vol. 55 (1944), pp. 68-84.

3. ———, "Intersections of algebraic and algebroid varieties," *ibid.*, vol. 57 (1945), pp. 1-85.

4. Wei-Liang Chow, "The Jacobian variety of an algebraic curve," to appear in this JOURNAL soon.

5. ——— and B. L. van der Waerden, "Zur algebraischen Geometrie. IX. Ueber zugeordnete Formen und algebraische Systeme von algebraischen Mannigfaltigkeiten," *Mathematische Annalen*, Band 113 (1937), pp. 692-704.

6. I. S. Cohen, "On the structure and ideal theory of complete local rings," *Transactions of the American Mathematical Society*, vol. 59 (1946), pp. 54-106.

7. W. Krull, "Dimensionstheorie in Stellenringen," *Journal für die reine und angewandte Mathematik*, vol. 179 (1938), pp. 204-226.

8. A. Weil, *Foundations of algebraic geometry*, New York, 1946.

9. O. Zariski, "Generalized semi-local rings," *Summa Brasiliensis Mathematicae*, vol. I, fasc. 8 (1946), pp. 169-195.

10. ———, "The concept of a simple point of an abstract algebraic variety," *Transactions of the American Mathematical Society*, vol. 62 (1947), pp. 1-52.

ARITHMETICAL PROPERTIES OF THE ELLIPTIC POLYNOMIALS ARISING FROM THE REAL MULTIPLICATION OF THE JACOBI FUNCTIONS.*

By MORGAN WARD.

I. Introduction.

1. In a previous paper in this JOURNAL,[1] referred to hereafter as "M," I have made a detailed investigation of the arithmetical properties of the sequence of polynomials (ψ),

$$\psi_n = \psi_n(\wp(u); g_2, g_3), \qquad (n = 0, 1, 2, \cdots)$$

associated with the real multiplication of the Weierstrass $\wp$ function when $\wp(u)$, g_2 and g_3 are given fixed rational values. If the elliptic discriminant $g_2{}^3 - 27g_3{}^2$ vanishes, (ψ) reduces essentially to Lucas' well-known linear sequence (U),

$$U_n = (\alpha^n - \beta^n)/(\alpha - \beta), \qquad (n = 0, 1, 2, \cdots).$$

I study here the arithmetical properties of the four polynomials A_n, B_n, C_n, and D_n associated with the real multiplication of Jacobi's *sn*, *cn*, and *dn*.[2]

Here each of $A_n, \cdots, D_n$ is a polynomial in sn^2u and k^2 with rational integral coefficients. Consequently, if we substitute for sn^2u and k^2 two fixed algebraic numbers, we obtain four sequences of algebraic numbers (A), (B), (C) and (D). The arithmetical properties of these elliptic sequences are the subject of this investigation. If k^2 is zero or one, the four sequences reduce essentially to Lucas' sequences (U) and (V), where $V_n = \alpha^n + \beta^n$.

2. To give an idea of the type of results obtained, choose fixed rational integral values x_0 and a_0 for sn^2u and k^2. Then the four elliptic sequences consist exclusively of rational integers. Each sequence is numerically periodic modulo m for any modulus m, but only the sequence (B) is an elliptic

---

* Received January 28, 1949.

[1] See the reference Ward [1] at the close of this paper.

[2] If n is an odd integer

$$snnu/snu = B_n/A_n, \quad cnnu/cnu = C_n/A_n, \quad dnu = D_n/A_n$$

with similar formulas when n is even.

divisibility sequence in the sense of M. The only primes whose laws of apparition present new features of interest are those dividing neither $2a_0(1-a_0)$ nor $x_0(1-x_0)(1-a_0x_0)$. Let p be such a prime, and let its rank of apparition in the divisibility sequence (B) be ρ, so that $B_n \equiv 0 \pmod{p}$ if and only if $n \equiv 0 \pmod{\rho}$. Then if ρ is odd, p divides no term of (A), (C) or (D). But if ρ is even, it appears as a divisor of precisely one of the sequences (A), (C) and (D). Suppose for example that it is a divisor of (C). Then no term of (A) or (D) is divisible by p, and $C_n \equiv 0 \pmod{p}$ if and only if n is an odd multiple of $\rho/2$.

The laws of repetition and apparition for powers of primes in (A), (B), (C) and (D) are easily reduced to the corresponding laws for (B) which in turn are corollaries of the results in Ward [2] for elliptic divisibility sequences.

3. The plan of the paper is sufficiently clear from the chapter titles. Accounts of the elliptic polynomials $A_n, \cdots, D_n$, are given in Krause [1] and Fricke [2]; but there are errors in the formulas given in these works. Although the properties of the *Al* functions[3] on which we base the theory were very completely worked out in Weierstrass [1], most of the formulas which we utilize are most simply obtained by transformation from the corresponding σ or θ function formulas.

II. Properties of Weierstrass *AL* Functions.

4. The multiplication theory of the Jacobian elliptic functions is most conveniently developed in terms of certain modified θ functions, the *Al* functions of Weierstrass (Weierstrass, [1]). These may be defined as follows: Let v be a complex variable, $q = e^{\pi i \tau}$ with $\operatorname{Im} \tau > 0$, $u = 2Kv$, where K is the complete elliptic integral. The Jacobi theta functions are then

$$\Theta(u) = \theta_0(v) = \sum_{-a}^{+a} (-1)^m q^{m^2} e^{2m\pi i v},$$

$$\mathrm{H}(u) = \theta_1(v) = -i \sum_{-a}^{+a} (-1)^m q^{(m+\frac{1}{2})^2} e^{(2m+1)\pi i v}$$

$$\mathrm{H}_1(u) = \theta_2(v) = \sum_{-a}^{+a} q^{(m+\frac{1}{2})^2} e^{(2m+1)\pi i v},$$

$$\Theta_1(u) = \theta_3(v) = \sum_{-a}^{+a} q^{m^2} e^{2m\pi i v}.$$

---

[3] The functions were named by Weierstrass in honor of Abel who was the first to consider them.

If we write θ_α for $\theta_\alpha(0)$ and θ_0 for $\theta_0(0)$, then Weierstrass Al functions may be defined by

$$(4.1)\qquad \begin{aligned} Al_1(u) &= \theta_3/\theta_0\theta_2 \exp(-\tfrac{1}{2}v^2\theta''_0/\theta_0)\theta_1(v), \\ Al_\alpha(u) &= 1/\theta_\alpha \exp(-\tfrac{1}{2}v^2\theta''_0/\theta_0)\theta_\alpha(v), \qquad (\alpha = 0, 2, 3). \end{aligned}$$

Note that $Al_1(u)$ is odd and $Al_2(u)$, $Al_3(u)$ and $Al_0(u)$ even. Also

$$(4.11)\qquad Al_1(0) = 0, \qquad Al_\alpha(0) = 1, \qquad (\alpha = 0, 2, 3).$$

sn, *cn* and *dn* have particularly simple expressions in terms of the Als; namely

$$(4.2)\qquad snu = Al_1(u)/Al_0(u),\ cnu = Al_2(u)/Al_0(u),\ dnu = Al_3(u)/Al_0(u).$$

The relationship to the Weierstrass σ functions is also very simple; namely if $w = \omega u/K$, then

$$(4.3)\qquad \begin{aligned} Al_1(u) &= (e_1 - e_3)^{\frac{1}{2}} e^{e_3 w^2/2}\sigma(w); \qquad & Al_3(u) &= e^{e_3 w^2/2}\sigma_2(w); \\ Al_2(u) &= e^{e_3 w^2/2}\sigma_1(w); & Al_0(u) &= e^{e_3 w^2/2}\sigma_3(w). \end{aligned}$$

For the lemniscate case, $e_3 = 0$ and the Al functions are essentially the σ functions.

The fundamental three-terms sigma identity becomes

$$(4.4)\qquad \begin{aligned} & Al_1(u + u_1)Al_1(u - u_1)Al_1(u_2 + u_3)Al_1(u_2 - u_3) \\ & + Al_1(u + u_2)Al_1(u - u_2)Al_1(u_3 + u_1)Al_1(u_3 - u_1) \\ & + Al_1(u + u_3)Al_1(u - u_3)Al_1(u_1 + u_2)Al_1(u_1 - u_2) = 0. \end{aligned}$$

5. We next introduce four new functions $K_{\alpha n}(u)$ of u and n by the definition

$$(5.1)\qquad K_{\alpha n} = Al_\alpha(nu)/Al_0(u)^{n^2}, \qquad (n = 0, 1, 2, \cdots;\ \alpha = 0, 1, 2, 3).$$

Evidently

$$(5.2)\qquad snnu = K_{1n}/K_{0n}, \qquad cnnu = K_{2n}/K_{0n}, \qquad dnnu = K_{3n}/K_{0n}.$$

The first three initial values of the four sequences (K_α) are as follows:

TABLE I. Initial values of $K_{\alpha n}$.

α/n	0	1	2
0	1	1	$1 - k^2sn^4u$.
1	0	snu	$2snu\ cnu\ dnu$.
2	1	cnu	$1 - 2sn^2u + k^2sn^4u$.
3	1	dnu	$1 - 2k^2sn^2n + k^2sn^4u$.

There are twenty-four addition formulas for the products $Al_\alpha(u+v) \times Al_\beta(u-v)$ obtainable by simple transformations from the corresponding formulas for $\theta_\alpha(u+v)\theta_\beta(u-v)$ or by suitably specializing (4.4). If in these formulas we replace u by nu and v by mu and divide by $Al_0(u)^{2(n^2+m^2)}$, we obtain on using (5.1) twenty-four addition formulas for the products $K_{\alpha n+m}K_{\beta n-m}$. It is sufficient to quote here a few such formulas as examples:

TABLE II. Addition formulas for $K_{\alpha n}$.

$$\text{(i)} \qquad K_{0n+m}K_{0n-m} = K^2{}_{0n}K^2{}_{0m} - k^2K^4{}_{1n}K^2{}_{1m}.$$

. . . . . .

$$\text{(v)} \qquad K_{1n+m}K_{1n-m} = K^2{}_{1n}K^2{}_{0m} - K^2{}_{0n}K^2{}_{1m}.$$

. . . . . .

$$\text{(xix)} \qquad K_{0n+m}K_{1n-m} = K_{0n}K_{1n}K_{2m}K_{3m} - K_{2n}K_{3n}K_{0m}K_{1m}.$$

. . . . . .

$$\text{(xxiv)} \qquad K_{2n+m}K_{3n-m} = K_{2n}K_{3n}K_{2m}K_{3m} - k'^2K_{0n}K_{1n}K_{0m}K_{1m}.$$

In formula (xxiv), k'^2 is the complementary modulus $1-k^2$.

If we take $m = \pm n$ or $m = n$ and $n = n+1$ in the addition formulas, we obtain a set of over forty " duplication formulas " which it is also unnecessary to give in detail; the two formulas so obtained from (i) and (xix) suffice as examples:

$$(5.3) \qquad K_{12n+1}K_{11} = K^2{}_{1n+1}K^2{}_{0n} - K^2{}_{0n+1}K^2{}_{1n},$$

$$(5.4) \qquad K_{12n} = 2K_{0n}K_{1n}K_{2n}K_{3n}.$$

As noted in Krause [1], the duplication formulas allow many results about the algebraic form of the polynomials $K_{\alpha n}$ to be proved by mathematical induction from the initial values given in Table I; the results in the next section are easily obtained in this manner.

III. The Four Elliptic Polynomials.

6. If we write

$$(6.1) \qquad \begin{array}{lll} K_{0n}(u) = A_n(sn^2u;k^2) & & n \text{ odd or even;} \\ & snuB_n(sn^2u;k^2) & n \text{ odd;} \\ K_{1n}(u) = & & \\ & snu\ cnu\ dnuB_n(sn^2u;k^2) & n \text{ even;} \\ & cnuC_n(sn^2u;k^2) & n \text{ odd;} \end{array}$$

$$K_{2n}(u) = \begin{cases} C_n(sn^2u; k^2) & n \text{ even}; \\ dnu D_n(sn^2u; k^2) & n \text{ odd}; \end{cases}$$

$$K_{3n}(u) = \begin{cases} D_n(sn^2u; k^2) & n \text{ even}; \end{cases}$$

then (Krause [1], pp. 159-162, Fricke [2], Chapter 2) A_n, B_n, C_n and D_n are polynomials in sn^2u and k^2 with rational integral coefficients. It is convenient to let

$$x = sn^2u, \qquad a = k^2. \tag{6.2}$$

Then

$$K_{1n}(u) = \begin{cases} (x)^{\frac{1}{2}} B_n(x; a), & n \text{ odd}; \\ (x(1-x)(1-ax))^{\frac{1}{2}} B_n(x; a), & n \text{ even}; \end{cases} \tag{6.3}$$

with similar formulas for K_{0n}, K_{2n} and K_{3n}.

We shall refer to A_n, B_n, C_n and D_n as the "elliptic polynomials of order n." If we let

$$\alpha_n = \begin{cases} (n^2-1)/2, & n \text{ odd}, \\ n^2/2, & n \text{ even}, \end{cases} \qquad \beta_n = \begin{cases} (n^2-1)/2 & n \text{ odd}, \\ (n^2-4)/2 & n \text{ even}, \end{cases} \tag{6.4}$$

then A_n, C_n and D_n are of degree α_n in x and B_n is of degree β_n in x.

7. There are a number of transformation formulas for the elliptic polynomials (Krause [1], Chapter III, Fricke [1], [2]) which are of arithmetical importance. These arise either by increasing u in the Al and K_n functions by the quarter periods K, iK' and $K + iK'$ or by performing the fundamental substitutions $\tau \to \tau + 1$, $\tau \to -1/\tau$ of the modular group on the four Al functions. It suffices here to develop one formula of each type by way of example.

Weierstrass showed that

$$Al_0(u+K) = 1/k'^{\frac{1}{2}} e^{\lambda(u^2-(u+K)^2)} Al_3(u),$$

$$Al_3(u+2K) = e^{\lambda(u^2-(u+K)^2)} Al_3(u),$$

where we have written λ for $(K-E)/2K$.

It easily follows that if n is odd,

$$Al_0(u+nK) = 1/k'^{\frac{1}{2}} e^{\lambda(u^2-(u+nK)^2)} Al_3(u).$$

Hence

$$K_{0n}(u+K)$$
$$=Al_0(nu+nK)/Al_0(u+K)^{n^2}=k'^{(n^2-1)/2}(Al_3(nu)/Al_3(u)^{n^2}.$$

On multiplying both sides of this expression by $dnu^{n^2}=(Al_3(u)/Al_0(u))^{n^2}$, we find that

$$(7.1) \qquad dnu^{n^2}K_{0n}(u+K)=k'^{(n^2-1)/2}K_{3n}(u).$$

Now $sn(u+K)=cnu/dnu$. Hence the substitution of $u+K$ for u induces the substitution of $(1-x)/(1-ax)$ for $x=sn^2u$. Therefore we obtain from (7.2) on substituting for K_{an} their expressions in terms of A_n and D_n the transformation formula

$$(1-ax)^{\alpha_n}A_n((1-x)/(1-ax))=(1-a)^{\alpha n/2}D_n(x) \qquad n \text{ odd}.$$

The following sets of transformation formulas are obtained by proceeding systematically in this manner.

8. The modular transformation $\tau\rightarrow-1/\tau$ is simply Jacobi's imaginary transformation $u\rightarrow iu$, $k\rightarrow k'$. Now Weierstrass[4] showed that

$$Al_3(iu;k')=e^{u^2/2}Al_3(u;k)$$
$$Al_0(iu;k')=e^{u^2/2}Al_2(u;k)$$

Hence

$$K_{3n}(iu;k')=Al_3(niu;k')/Al_0(iu;k')^{n^2}=Al_3(nu;k)/Al_2(u;k)^{n^2},$$

or

$$(8.1) \qquad K_{3n}(iu;k')=cnu^{-n^2}K_{3n}(u;k).$$

But since $sn(iu,k')=isn(u,k)/cn(u,k)$, Jacobi's imaginary transformation induces the transformation

$$(8.2) \qquad x\rightarrow -x/(1-x), \qquad a\rightarrow 1-a$$

on x and a.

Now $K_{3n}=(1-ax)^{\frac{1}{2}}D_n$ or D_n according as n is odd or even, and (8.2) throws $\sqrt{1-ax}$ into $(1-ax)^{\frac{1}{2}}/1-x$. Hence on substituting into (8.1) and using the abbreviation α_n for $(n^2-1)/2$ or $n^2/2$ we obtain the formula

$$(1-x)^{\alpha_n}D_n(x/(x-1);1-a)=D_n(x;a).$$

The formulas listed below were obtained by systematically combining

---

[4] Weierstrass [1], page 20.

TABLE III. Transformation formulas.

n odd	*n even*
$(1-ax)^{\alpha_n}A_n((1-x)/(1-ax)) = (1-a)^{\alpha_{n/2}}D_n(x)$;	$= (1-a)^{\alpha_{n/2}}A_n(x)$
$(1-ax)^{\beta_n}B_n((1-x)/(1-ax)) = (-1)^{(n-1)/2}(1-a)^{\beta_{n/2}}C_n(x)$;	$= (-1)^{(n-2)/2}(1-a)^{\beta_{n/2}}(x)$
$(1-ax)^{\alpha_n}C_n((1-x)/(1-ax)) = (-1)^{(n-1)/2}(1-a)^{\alpha_{n/2}}B_n(x)$;	$= (-1)^{n/2}(1-a)^{\alpha_{n/2}}C_n(x)$
$(1-ax)^{\alpha_n}D_n(1-x)/(1-ax)) = (1-a)^{\alpha_{n/2}}A_n(x)$;	$= (1-a)^{\alpha_{n/2}}D_n(x)$
$(\sqrt{ax})^{\alpha_n}A_n(1/ax) = (-1)^{(n-1)/2}B_n(x)$;	$= (-1)^{n/2}A_n(x)$
$(\sqrt{ax})^{\beta_n}B_n(1/ax) = (-1)^{(n-1)/2}A_n(x)$;	$= (-1)^{(n-2)/2}B_n(x)$
$(\sqrt{ax})^{\alpha_n}C_n(1/ax) = D_n(x)$;	$= C_n(x)$
$(\sqrt{ax})^{\alpha_n}D_n(1/ax) = C_n(x)$;	$= D_n(x)$
$(1-x)^{\alpha_n}A_n((1-ax)/(a-ax)) = ((1-a)/a)^{\alpha_{n/2}}C_n(x)$;	$= (-1)^{n/2}((1-a)/a)^{\alpha_{n/2}}A_n(x)$
$(1-x)^{\beta_n}B_n((1-ax)/(a-ax)) = (-1)^{(n-1)/2}((1-a)/a)^{\beta_{n/2}}D_n(x)$;	$= ((1-a)/a)^{\beta_{n/2}}B_n(x)$
$(1-x)^{\alpha_n}C_n((1-ax)/(a-ax)) = ((1-a)/a)^{\alpha_{n/2}}A_n(x)$;	$= (-1)^{n/2}((1-a)/a)^{\alpha_{n/2}}C_n(x)$
$(1-x)^{\alpha_n}D_n((1-ax)/(a-ax)) = (-1)^{(n-1)/2}((1-a)/a)^{\alpha_{n/2}}B_n(x)$;	$= ((1-a)/a)^{\alpha_{n/2}}D_n(x)$.

the two transformations $\tau \to -1/\tau$ and $\tau \to 1 + \tau$; the latter transformation induces the transformation

$$(8.3) \qquad x \to (x - ax)/(1 - ax), \qquad a \to a/(a-1)$$

on x and a.

TABLE IV. Transformation formulas.

$$\begin{aligned} A_n(x;a) &= (1-ax)^{\alpha_n} D_n((x-ax)/(1-ax);a/(a-1)) \\ &= (1-x)^{\alpha_n} C_n(x/(x-1);1-a), \\ &= (1-ax)^{\alpha_n} C_n(ax/(ax-1);(a-1)/a) = A_n(ax;1/a) \\ &= (1-x)^{\alpha_n} D_n((ax-x)/(1-x);1/(1-a). \end{aligned}$$

$$\begin{aligned} B_n(x;a) &= (1-ax)^{\beta_n} B_n((x-ax)/(1-ax);a/(a-1)) \\ &= (1-x)^{\beta_n} B_n(x/(x-1);1-a), \\ &= (1-ax)^{\alpha_n} B_n(ax/(ax-1);(a-1)/a) = B_n(ax;1/a) \\ &= (1-x)^{\beta_n} B_n((ax-x)/(1-x);1/(1-a). \end{aligned}$$

$$(8.4) \quad \begin{aligned} C_n(x;a) &= (1-ax)^{\alpha_n} C_n((x-ax)/(1-ax);a/(a-1)) \\ &= (1-x)^{\alpha_n} A_n(x/(x-1);1-a), \\ &= (1-ax)^{\alpha_n} D_n(ax/(ax-1);(a-1)/a) = D_n(ax;1/a) \\ &= (1-x)^{\alpha_n} A_n((ax-x)/(1-x);1/(1-a)). \end{aligned}$$

$$\begin{aligned} D_n(x;a) &= (1-ax)^{\alpha_n} A_n((x-ax)/(1-ax);a/(a-1)) \\ &= (1-x)^{\alpha_n} D_n(x/(x-1);1-a), \\ &= (1-ax)^{\alpha_n} A_n(ax/(ax-1);(a-1)/a) = C_n(ax;1/a) \\ &= (1-x)^{\alpha_n} C_n((ax-x)/(1-x);1/(1-a)). \end{aligned}$$

We finally tabulate for later reference the first few initial values of the four elliptic sequences.

TABLE V. Initial values.

n	0	1	2	3
A_n	1	1	$1 - ax^2$	$1 - 6ax^2 + 4a(1+a)x^3 - 3a^2x^4$
B_n	0	1	2	$3 - 4(1+a)x + 6ax^2 - a^2x^4$
C_n	1	1	$1 - 2x + ax^2$	$1 - 4x + 6ax^2 - 4a^2x^3 + a^2x^4$
D_n	1	1	$1 - 2ax + ax^2$	$1 - 4ax + 6ax^2 - 4ax^3 + a^2x^4$.

The transformation formulas of Tables III and IV may all be proved without function theory by mathematical induction from the duplication formulas and the initial values in Table V.

IV. Elliptic Divisibility Sequences in Domains of Integrity.

9. In M, the functional equation

$$\omega_{m+n}\omega_{m-n} = \omega_{m+1}\omega_{m-1}\omega^2_n - \omega_{n+1}\omega_{n-1}\omega^2_m \tag{9.1}$$

was solved completely over the ring of rational integers and over the field of all complex numbers. It is necessary for the purposes of this paper to extend certain theorems of M to solutions of (9.1) over more general rings.

Let $\mathfrak{R}$ denote a domain of integrity; (commutative ring with a unit and no divisors of zero). $\mathfrak{R}$ may be a field; in any event we denote its quotient field by $\mathfrak{F}$, and for brevity refer to $\mathfrak{R}$ as a ring. We are interested in solutions of (9.1) over $\mathfrak{R}$ and over $\mathfrak{F}$. We lay down the following definitions:

A particular solution

$$(h): \quad h_0, h_1, h_2, \cdots, h_n, \cdots$$

of (9.1) will be said to belong to $\mathfrak{R}$ (to $\mathfrak{F}$) if all its terms belong to $\mathfrak{R}$ (to $\mathfrak{F}$). If a and b belong to $\mathfrak{R}$, we say that a divides b in $\mathfrak{R}$ if there exists an element c of $\mathfrak{R}$ such that $ac = b$. We write: $a \mid b\ (\mathfrak{R})$.

In particular, if $\mathfrak{R}$ is a field, and $b \neq 0$, $a \mid b\ (\mathfrak{R})$ for every $a \neq 0$.

If $\mathfrak{m}$ is an ideal of $\mathfrak{R}$, $a \equiv b\ (\mathfrak{m})$ means as usual $a - b$ is contained in $\mathfrak{m}$.

If $\mathfrak{p}$ is a maximal ideal of $\mathfrak{R}$, the quotient ring $\mathfrak{R}/\mathfrak{p}$ is a field. If this field is finite its order is a power of a certain rational prime p. We denote the order by $N\mathfrak{p} = p^f$ and call p the (rational) prime belonging to $\mathfrak{p}$.

Definition 9.1. *A solution of* (9.1) *is said to be "regular over* $\mathfrak{F}$*" if it belongs to* $\mathfrak{F}$ *and if*

$$h_0 = 0, \qquad h_1 = 1, \qquad h_2, h_3 \text{ not both zero.} \tag{9.2}$$

Definition 9.2. *A solution* (h) *of* (9.1) *is said to be "a divisibility sequence over* $\mathfrak{R}$*" if it belongs to* $\mathfrak{R}$ *and if*

$$h_r \mid h_s\ (\mathfrak{R}) \qquad \text{whenever } r \mid s. \tag{9.3}$$

Let (h) belong to $\mathfrak{R}$, and let $\mathfrak{m}$ be an ideal of $\mathfrak{R}$. $\mathfrak{m}$ is called a divisor of (h) if it contains at least one term h_{n_0} of the sequence (h) with $n_0 > 0$. n_0 is called a "place of apparition" of $\mathfrak{m}$ in (h). If in addition $h_r \not\equiv 0$ (mod $\mathfrak{m}$) for every proper divisor r of n_0, then n_0 is called a rank of apparition of $\mathfrak{m}$ in (h).

10. The proofs of the theorems which follow are by mathematical induction with the exception of the proof of Theorem 10.7 which uses the Dirichlet box principle. In any event they are almost word for word the same as corresponding theorems in M for the special cases when $\mathfrak{R}$ is the ring of rationals or the complex field. We shall accordingly cite the corresponding results in M for the details of proof.

THEOREM 10.1. *Let* (h) *be a sequence satisfying the following three conditions:*

$$(10.1) \qquad (h) \text{ is a regular solution of } (9.1) \text{ over } \mathfrak{F},$$

$$(10.2) \qquad h_2, h_3 \text{ and } h_4 \text{ belong to } \mathfrak{R},$$

$$(10.3) \qquad h_2 \mid h_4\ (\mathfrak{R}).$$

Then (h) *is a divisibility sequence over* $\mathfrak{R}$ *uniquely determined by its initial values* h_2, h_3 *and* h_4.

Proof. M, Theorem 4.1. Chapter II.

THEOREM 10.2. *Under the hypotheses of Theorem* 10.1,

$$(10.4) \qquad h_n = P_n(h_2, h_3, h_4),$$

where P_n *is a polynomial in* h_2, h_3, h_4 *with rational integral coefficients and such that for any element* a *of* $\mathfrak{F}$

$$(10.5) \qquad P_n(a^3h_2, a^8h_3, a^{15}h_4) = a^{n^2-1}P_n(h_2, h_3, h_4).$$

Proof. M, Theorem 4.1. Chapter II.

THEOREM 10.3. *If* (k) *is any regular solution of* (9.1) *over* $\mathfrak{F}$ *and* a *any non-zero element of* $\mathfrak{F}$, *then* (l) *is also a regular solution over* $\mathfrak{F}$, *where*

$$l_n = a^{n^2-1}k_n, \qquad (n = 0, 1, 2, \cdots).$$

THEOREM 10.4. *If* (k) *is any regular solution of* (1.1) *over* $\mathfrak{F}$, *then there always exists an element* a *of* $\mathfrak{F}$ *and a regular solution* (h) *of* (9.1) *over* $\mathfrak{R}$ *satisfying conditions* (10.1), (10.2) *and* (10.3) *such that*

$$k_n = a^{n^2-1}h_n, \qquad (n = 0, 1, 2, \cdots).$$

Proof. M, Theorems 21.1, 21.3.

THEOREM 10.5. *Let* (k) *be a regular solution of* (1.1) *over* $\mathfrak{F}$ *with* $k_2 \neq 0$. *Then if two consecutive terms of* (k) *are zero, all terms vanish beyond the third.*

THEOREM 10. 6. *Let* (h) *be a solution of* (1. 1) *over* $\mathfrak{R}$ *with* $h_0 = 0$ *and* $h_1 = 1$ *(but not necessarily a regular solution). Then if* (h) *is a divisibility sequence over* $\mathfrak{R}$ *and two consecutive terms of* (h) *vanish, all terms of* (h) *vanish beyond the third.*

Proof. M, Lemma 4. 1.

It is shown in M that the hypothesis that (h) is a divisibility sequence in Theorem 10. 6 is necessary for the truth of the theorem when (h) is not regular; that is, when both h_2 and h_3 are zero.

THEOREM 10. 7. *Let* (h) *be an elliptic divisibility sequence over* $\mathfrak{R}$, *and let* $\mathfrak{p}$ *be a prime ideal of* $\mathfrak{R}$ *whose quotient ring* $\mathfrak{R}/\mathfrak{p}$ *is of finite order* $N\mathfrak{p}$. *Then* $\mathfrak{p}$ *is a divisor of* (h), *and has a rank of apparition* r *in* (h) *less than* $2(N\mathfrak{p}+1)$.

Proof. M, Theorem 5. 1.

It is shown in M that the upper limit $2(N\mathfrak{p}+1)$ is the best possible for the rank of $\mathfrak{p}$ in (h).

V. The Laws of Apparition of Prime Ideals in Elliptic Sequences.

11. The connection between the results of Chapter IV and the elliptic polynomials is made by the following theorem.

THEOREM 11. 1. *If* u_0 *is neither a zero nor a pole of* snu *and if*

$$h_n = K_{1n}(u_0)/K_{11}(u_0) = K_{1n}/K_{11} \qquad (n = 0, 1, 2, \cdots) \tag{11. 1}$$

then the sequence (h) *is a solution of the functional equation*

$$\omega_{m+n}\omega_{m-n} = \omega_{m+1}\omega_{m-1}\omega^2_n - \omega_{n+1}\omega_{n-1}\omega^2_m \tag{9. 1}$$

over the complex field.

Proof. Let l, m and n be fixed integers. Take $u = 0$, $u_1 = lu_0$, $u_2 = mu_0$ and $u_3 = nu_0$ in the basic three-term identity (4. 4), and divide by the non-zero quantity $Al_0(u_0)^{2(l^2+m^2+n^2)}$. Then we obtain on substitution, from (5. 1) the formula

$$K^2_{1l}K_{1m+n}K_{1m-n} + K^2_{1m}K_{1n+l}K_{1n-l} + K^2_{1n}K_{1l+m}K_{1l-m} = 0.$$

Now $K_{11} = snu_0 \neq 0$. Hence on letting $l = 1$ and dividing by K^4_{11}, we obtain from (11. 1) the formula

$$h_{m+n}h_{m-n} = h_{m+1}h_{m-1}h^2_n - h_{n+1}h_{n-1}h^2_m;$$

for $K_{1l-m} = -K_{1m-l}$.

We shall now assign to x and to a algebraic integer values x_0 and a_0. Let $\mathfrak{F}_1$ be the field $F[x_0, a_0]$ obtained by adjoining x_0 and a_0 to the rational field F, and let $\mathfrak{R}_1$ be the ring of integers of $\mathfrak{F}_1$. Let $\mathfrak{F}$ be the field $F[x_0^{\frac{1}{2}}, 1-x_0^{\frac{1}{2}}, 1-a_0x_0^{\frac{1}{2}}]$, and $\mathfrak{R}$ the ring of integers of $\mathfrak{F}$. Clearly $\mathfrak{R} \supseteq \mathfrak{R}_1$, $\mathfrak{F} \supseteq \mathfrak{F}_1$, and every ideal of either ring has a finite norm. We shall use German letters to denote either ideals of $\mathfrak{R}$ or of $\mathfrak{R}_1$, and let $\mathfrak{p}$ denote as usual a prime ideal, and p the corresponding rational prime.

The four elliptic sequences (A), (B), (C) and (D) belong to $\mathfrak{R}_1$ while the four sequences (K_α), $(\alpha = 0, 1, 2, 3)$ belong to $\mathfrak{R}$.

Let u_0 be chosen in a period parallelogram so that[5]

$$snu_0 = (x_0)^{\frac{1}{2}}.$$

Then

$$(11.2) \qquad h_n = \begin{cases} B_n(x_0, a_0), & n \text{ odd}, \\ ((1-x_0)(1-a_0x_0))^{\frac{1}{2}} B_n(x_0, a_0), & n \text{ even}. \end{cases}$$

Hence the initial values of (h) are

$$h_0 = 0, \qquad h_1 = 1, \qquad h_2 = 2((1-x_0)(1-a_0x_0))^{\frac{1}{2}},$$
$$h_3 = B_3(x_0, a_0), \qquad h_4 = ((1-x_0)(1-a_0x_0))^{\frac{1}{2}} B_4(x_0, a_0).$$

Now by direct calculation or from Table III,

$$(11.3) \qquad B_3(1; a_0) = -(1-a_0)^2, \qquad B_3(1/a_0; a_0) = -(1-a_0/a_0^2)^2.$$

Hence h_2 and h_3 both vanish if and only if $x_0 = 1$ and $a_0 = 1$. Also $h_4 = 2h_2A_2C_2D_2$ by formula (5.4). Hence $h_2 \mid h_4\ (\mathfrak{R})$.

Thus the sequence (h) satisfies all the hypotheses of Theorem 10.1. We may consequently state

THEOREM 11.1. *Unless both x_0 and a_0 are unity, the sequence (h) defined by*

$$(11.2) \qquad h_n = B_n(x_0; a_0), \ n \text{ odd}; \qquad = ((1-x_0)(1-a_0x_0))^{\frac{1}{2}} B_n(x_0; a_0), \ n \text{ even}$$

is an elliptic divisibility sequence over the ring $\mathfrak{R}$. Every prime ideal $\mathfrak{p}$ of $\mathfrak{R}$ is a divisor of (h). Furthermore, if $\mathfrak{p}$ does not divide both h_3 and h_4 then $\mathfrak{p}$ has a unique rank of apparition ρ in (h) such that

$$(11.4) \qquad h_n \equiv 0 \ (\mathrm{mod}\ \mathfrak{p}) \qquad \textit{if and only if } n \equiv 0 \ (\mathrm{mod}\ \rho).$$

---

[5] Two choices of u_0 are possible; for definiteness let u_0 be chosen with smallest imaginary part; if u_0 is real, with smallest real part.

If $\mathfrak{p}$ does divide both h_3 and h_4, it divides every subsequent term of (h) by Theorem 10.5. We call such primes "null divisors" of (h).

12. Let $\mathfrak{p}$ be a null divisor of (h). Then

$$(12.1)\quad B_3(x_0;a_0) \equiv 0 \pmod{\mathfrak{p}},$$
$$((1-x_0)(1-a_0x_0))^{\frac{1}{2}}B_4(x_0;a_0) \equiv 0 \pmod{\mathfrak{p}}.$$

We first consider the case when $\mathfrak{p}$ divides both B_3 and B_4. Then the two polynomials $B_3(x, a_0)$ and $B_4(x, a_0)$ have a common root in the field $\mathfrak{R}_1/\mathfrak{p}$. Hence their resultant must vanish in this field. Now this resultant is found to have the value $2^{20}a_0{}^4(1-a_0)^9$. Hence either $a_0 \equiv 0$, $a_0 \equiv 1$ or $2 \equiv 0 \pmod{\mathfrak{p}}$.

$a_0 \not\equiv 0 \pmod{\mathfrak{p}}$. For if $a_0 \equiv 0$, then by Table V, $B_3(x_0, a_0) \equiv 3 - 4x_0$ and $B_4(x_0, a_0) \equiv 4 - 8x_0 \pmod{\mathfrak{p}}$. But the congruences $3 - 4x_0 \equiv 4 - 8x_0 \equiv 0 \bmod \mathfrak{p})$ are impossible.

If $a_0 \equiv 1 \pmod{\mathfrak{p}}$, *then* $x_0 \equiv 1 \pmod{\mathfrak{p}}$. For by Tables IV and V,

$$\left.\begin{aligned} B_3(x_0, a_0) &\equiv (1-x_0)^3(3-x_0) \\ B_4(x_0, a_0) &\equiv 4(1-x_0)^5(1+x_0) \end{aligned}\right\} \pmod{\mathfrak{p}}$$

if $a_0 \equiv 1$. Hence if $x_0 \not\equiv 1 \pmod{\mathfrak{p}}$, we must have $3 - x_0 \equiv 2 + 2x_0 \equiv 0 \pmod{\mathfrak{p}}$ but $1 - x_0 \not\equiv 0 \pmod{\mathfrak{p}}$, which is impossible.

Now finally if $2 \equiv 0 \pmod{\mathfrak{p}}$ but $a_0 \not\equiv 1 \pmod{\mathfrak{p}}$, since $B_4(x_0;a_0) \equiv 0 \pmod{2}$ and $B_3(x_0;a_0) \equiv (1-a_0x_0{}^2)^2 \pmod{2}$ we must have $1 - a_0x_0{}^2 \equiv 0 \pmod{\mathfrak{p}}$. Hence since $a_0 \not\equiv 1 \pmod{\mathfrak{p}}$,

$$((1-x_0)(1-a_0x_0))^{\frac{1}{2}} \not\equiv 0 \pmod{\mathfrak{p}}.$$

If on the other hand $((1-x_0)(1-a_0x_0))^{\frac{1}{2}} \equiv 0 \pmod{\mathfrak{p}}$, it is easily shown that the previous case $a_0 \equiv x_0 \equiv 1 \pmod{\mathfrak{p}}$ must hold. We have thus proved

THEOREM 12.1. *The only prime ideal null divisors of (h) are primes dividing $2(1-a_0)$. Necessary and sufficient conditions that $\mathfrak{p}$ be a null divisor are that either*

$$a_0 \equiv 1 \text{ and } x_0 \equiv 1 \pmod{\mathfrak{p}} \text{ or}$$
$$2 \equiv 0 \text{ and } a_0x_0{}^2 \equiv 1 \pmod{\mathfrak{p}} \text{ but}$$
$$a_0 \not\equiv 1 \text{ and } x_0 \not\equiv 1 \pmod{\mathfrak{p}}.$$

In the lemniscate case, for example when $a_0 = -1$, the only null divisors are divisors of two for which $x_0 \equiv 1 \pmod{\mathfrak{p}}$. Hence if x_0 is an even rational integer, there are no null divisors.

13. We may classify the prime ideals of both $\mathfrak{R}$ and $\mathfrak{R}_1$ into three categories:

I. Ideals dividing $2a_0(1-a_0)$.
II. Ideals dividing $x_0(1-x_0)(1-a_0x_0)$.
III. Ideals dividing neither $2a_0(1-a_0)$ nor $x_0(1-x_0)(1-a_0x_0)$.

Ideals of the first and second categories will be called irregular; they include all null divisors, and are usually finite in number. Ideals of the third category will be called regular. In this section, we shall determine their laws of apparition in the elliptic sequences (A), (B), (C) and (D).

THEOREM 13. 1. *No regular prime ideal can divide any two of A_n, B_n, C_n and D_n for the same value of n. Common divisors of A_n, B_n; A_n, C_n; $\cdots$, C_n, D_n are always null divisors of the sequence (B).*

Proof. Suppose, for example, that for a certain fixed value of n

$$A_n \equiv 0 \pmod{\mathfrak{p}} \text{ and } B_n \equiv 0 \pmod{\mathfrak{p}}, \qquad \mathfrak{p} \text{ regular.}$$

Then $K_{0n} \equiv 0 \pmod{\mathfrak{p}}$ and $K_{1n} \equiv 0 \pmod{\mathfrak{p}}$. Hence by the duplication formulas (5. 3) and (5. 4),

$$K_{12n} \equiv K_{11}K_{12n+1} \equiv 0 \pmod{\mathfrak{p}}$$

so that $K_{11}h_{2n} \equiv K^2{}_{11}h_{2n+1} \equiv 0 \pmod{\mathfrak{p}}$ by formula (11. 1). But $K^2{}_{11} = sn^2u_0 = x_0 \not\equiv 0 \pmod{\mathfrak{p}}$. Hence $\mathfrak{p}$ divides two consecutive terms of the elliptic divisibility sequence (h). Therefore by Theorems 11. 1, 10. 5 and 12. 2, $h_3 \equiv h_4 \equiv 0 \pmod{\mathfrak{p}}$, $2(1-a_0) \equiv 0 \pmod{\mathfrak{p}}$, contrary to the hypothesis that $\mathfrak{p}$ is regular.

It can be shown from the duplication formulas that a similar contradiction ensues if it is assumed that any other pair from A_n, B_n, C_n and D_n is divisible by $\mathfrak{p}$.

14. For regular prime ideals, the rank ρ of $\mathfrak{p}$ in (h) and in (B) is evidently the same. Hence if $\mathfrak{p}$ is regular,

$$B_n \equiv 0 \pmod{\mathfrak{p}} \qquad \text{if and only if } n \equiv 0 \pmod{\rho}, \tag{14. 1}$$

where ρ is a fixed positive integer depending only on $\mathfrak{p}$ and B_3 and B_4.

THEOREM 14. 1. *Let $\mathfrak{p}$ be a regular prime ideal. Then if the rank of apparition of $\mathfrak{p}$ in (B) is odd, $\mathfrak{p}$ is not a divisor of (A), (C) or (D).*

Proof. Let $\mathfrak{p}$ be regular. Then $B_n \equiv 0 \pmod{\mathfrak{p}}$ if and only if $K_{1n} \equiv 0 \pmod{\mathfrak{p}}$; similarly A_n, C_n or D_n are divisible by $\mathfrak{p}$ only if K_{0n}, K_{2n} or K_{3n} are divisible by $\mathfrak{p}$. Suppose that $\mathfrak{p}$ is of odd rank ρ in (B) and a divisor of (A), for example. Then there exists a term A_k of (A) such that $A_k \equiv 0 \pmod{\mathfrak{p}}$. Now by the duplication formula (5.4),

$$K_{2k} \equiv 2K_{0k}K_{1k}K_{2k}K_{3k}. \tag{14.2}$$

Hence by the preceding remarks, $B_{2k} \equiv 0 \pmod{\mathfrak{p}}$ so that $\rho \mid 2k$ by (14.1). Therefore $\rho \mid k$ so that $A_k \equiv B_k \equiv 0 \pmod{\mathfrak{p}}$ contrary to Theorem 13.1. In like manner, $\mathfrak{p}$ cannot be a divisor of (C) or (D).

THEOREM 14.2. *Let $\mathfrak{p}$ be a regular prime ideal of even rank of apparition in (B). Then $\mathfrak{p}$ is a divisor of precisely one of the three sequences (A), (C) and (D). If $\mathfrak{p}$ is a divisor of (C), then $C_n \equiv 0 \pmod{\mathfrak{p}}$ if and only if n is an odd multiple of $\rho/2$, with similar results if $\mathfrak{p}$ is a divisor of (A) or (D).*

Proof. Let $\mathfrak{p}$ and ρ satisfy the hypothesis of the theorem. Then by the duplication formula (5.4)

$$K_{1\rho} \equiv 2K_{0(\rho/2)}K_{1(\rho/2)}K_{2(\rho/)}K_{3(\rho/2)} \pmod{\mathfrak{p}}.$$

Hence since $B_\rho \equiv 0 \pmod{\mathfrak{p}}$ and $\mathfrak{p}$ is regular, precisely one of $A_{\rho/2}$, $B_{\rho/2}$, $C_{\rho/2}$, $D_{\rho/2}$ is divisible by $\mathfrak{p}$. Evidently, $B_{\rho/2} \not\equiv 0 \pmod{\mathfrak{p}}$. Assume that $C_{\rho/2} \equiv 0 \pmod{\mathfrak{p}}$ so that $\mathfrak{p}$ is a divisor of (C). Then $\mathfrak{p}$ is not a divisor of (A) or (D). For if for example $D_k \equiv 0 \pmod{\mathfrak{p}}$, then by the duplication formula (14.2), $B_{2k} \equiv 0 \pmod{\mathfrak{p}}$. Hence $\rho \mid 2k$, $\rho/2 \mid k$ and $B_k \equiv 0 \pmod{\mathfrak{p}}$ contrary to Theorem 13.1. In precisely the same way we can show that $A_k \not\equiv 0 \pmod{\mathfrak{p}}$, and that if $C_k \equiv 0 \pmod{\mathfrak{p}}$, then k must be an odd multiple of $\rho/2$. It remains to prove the converse of this last statement. Consider then any term C_k of the sequence (C) in which k is a multiple of $\rho/2$. Then $2k$ is a multiple of ρ. Hence by (14.1) and (14.2)

$$0 \equiv 2A_kB_kC_kD_k \pmod{\mathfrak{p}}.$$

Now either B_k or C_k must be divisible by $\mathfrak{p}$ but not both, by what we have already proved. If k is an even multiple of $\rho/2$, it is a multiple of ρ, so that $B_k \equiv 0$ and $C_k \not\equiv 0 \pmod{\mathfrak{p}}$. But if k is an odd multiple of $\rho/2$, it is not a multiple of ρ. Consequently, $B_k \not\equiv 0 \pmod{\mathfrak{p}}$, so that $C_k \equiv 0 \pmod{\mathfrak{p}}$, completing the proof for regular divisors of (C). The proof for divisors of (A) or (D) is precisely similar.

15. It remains to discuss the laws of apparition of the irregular prime ideals of categories I and II in Section **13**. Since the elliptic polynomials have rational integral coefficients, if $\mathfrak{p}$ is a prime ideal dividing a_0 say, we have $A_n(x; a_0) \equiv A_n(x; 0) \pmod{\mathfrak{p}}$. Consequently the arithmetical behavior of the sequences modulo $\mathfrak{p}$ is given immediately by the algebraic behavior of the elliptic polynomials in the following five singular cases:

(15.1) (i) $a=0$; (ii) $a=1$; (iii) $x=0$; (iv) $x=1$; (v) $ax=1$.

Here the first two cases apply to all prime ideals of category I save divisors of two, while the last three cases apply to prime ideals of category II. We discuss the former two cases in this section, and the latter three in Section **16.**

Case 15.1 (i) $a=0$.

Then $k^2=0$ and snu becomes $\sin u$, cnu becomes $\cos u$ and dnu becomes one. Thus if

$$U_n=U_n(x)=(\alpha^n-\beta^n)/(\alpha-\beta) \qquad V_n=V_n(x)=\alpha^n+\beta^n$$

are the Lucas functions of the quadratic equation $t^2-2t\sqrt{1-x}+1$ where

$$x=\sin^2 u, \tag{15.2}$$

then we readily find that

$$\begin{aligned} A_n(x;0) &= D_n(x;0)=1; && \\ B_n(x;0) &= \sin nu/\sin u = U_n(x), && n \text{ odd},\\ &= \sin nu/(\sin u\cos u)=2U_n(x)/V_1(x), && n \text{ even};\\ C_n(x;0) &= \cos nu/\cos u = V_n(x)/V_1(x), && n \text{ odd}\\ &\cos nu = V_n(x)/2, && n \text{ even}. \end{aligned}$$

We thus obtain the following theorem:

Theorem 15.1. *If $\mathfrak{p}$ is a prime ideal of the first category dividing a_0, then*

$$\begin{aligned} &A_n \equiv D_n \equiv 1 \pmod{\mathfrak{p}}\\ &B_n \equiv U_n,\quad n \text{ odd};\quad \equiv 2U_n/V_1,\quad n \text{ even};\\ &C_n \equiv V_n/V_1,\quad n \text{ odd};\quad \equiv \tfrac{1}{2}V_n,\quad n \text{ even}. \end{aligned}$$

Here $U_n=U_n(x_0)$ and $V_n\equiv V_n(x_0)$ are the Lucas functions of the quadratic equation $t^2-2(1-x_0)^{\frac{1}{2}}t+1=0$.

Case 15.1 (ii). $a=1$.

Then $k^2=1$ and snu becomes $\tanh u$, while cnu and dnu become $\operatorname{sech} u$. Now by the transformation formulas of Table IV,

$$\begin{aligned} A_n(x;1) &= (1-x)^{\alpha_n}C_n(x/(x-1);0)\\ B_n(x;1) &= (1-x)^{\beta_n}B_n(x/(x-1);0)\\ C_n(x;1) &= (1-x)^{\alpha_n}A_n(x/(x-1);0)\\ D_n(x;1) &= (1-x)^{\alpha_n}D_n(x/(x-1);0). \end{aligned}$$

Hence if we let $u = ir$, then $x = -\tan^2 r$ and $x/(x-1) = \sin^2 r$. Therefore, by the results of case (i),

$$\begin{aligned} A_n(x;1) &= \sec r^{n^2} \cos nr = (2/V_1)^{n^2} V_n/2 \text{ [6]} \\ &= \sec r^{n^2-3} (\sin nr/\sin r) = (2/V_1)^{n^2-3} U_n, && n \text{ even} \\ B_n(x;1) &= \sec r^{n^2-1} (\sin nr/\sin r) = (2/V_1)^{n^2-1} U_n, && n \text{ odd} \\ C_n(x;1) &= D_n(x;1) = \sec r^{2a n} = (2/V_1)^{2a n}. \end{aligned}$$

Here $U_n = U_n(x/(x-1))\, V_n = V_n(x/(x-1))$ are the Lucas functions of the quadratic equation $t^2 - 2t/(1-x)^{\frac{1}{2}} + 1 = 0$.

We thus obtain the following theorem:

THEOREM 15.2. *If $\mathfrak{p}$ is a prime ideal of the first category dividing $1 - a_0$, then*

$$A_n \equiv (2/V_1)^{n^2} V_n/2);$$

$$B_n \equiv \begin{matrix} (2/V_1)^{n^2-1} U_n, & n \text{ odd}, \\ (2/V_1)^{n^2-3} U_n, & n \text{ even}; \end{matrix}$$

$$C_n \equiv D_n \equiv \begin{matrix} (2/V_1)^{n^2-1}, & n \text{ odd}, \\ (2/V_1)^{n^2}\ , & n \text{ even}. \end{matrix}$$

Here $U_n = U_n(x_0/(x-1))$, $V_n = V_n(x_0/(x_0-1))$ are the Lucas functions of the quadratic equation $t^2 - 2t/(1-x_0)^{\frac{1}{2}} + 1 = 0$.

For prime ideals of the first category dividing two, a special discussion must be made as in the case of the rational field for the prime two treated in M, pages 40-41. We shall not pursue the matter further here.

16. Consider now prime ideals of the second category. We may confine ourselves to ideals which are not also of the first category; for ideals of both categories are either null divisors of (B) or else are already covered by the results of Section **15**. A prime ideal of this character is easily shown to divide precisely one of the algebraic numbers x_0, $1 - x_0$ or $1 - a_0 x_0$. Clearly then $A_n(x_0, a_0)$ is congruent to either $A_n(0, a_0)$, $A_n(1, a_0)$ or $A_n(1/a_0, a_0)$, with similar results for B_n, C_n and D_n.

By the results of Table IV, we find that

---

[6] For example, $A_3(x;1) = (1 - x^3)(1 + 3x)$ by direct calculation from Table V. On the other hand, the formula for $n = 3$ becomes $A_3(x;1) = \sec r^8 (\cos 3r/\cos r)$. Now $\cos 3r/\cos r = 4\cos^2 r - 3 = [1 + 3(-\tan^2 r)]/\sec^2 r$ and $\sec^2 r = 1 - (-\tan^2 r)$. Hence $A_3(x;1) = (1 - (-\tan^2 r))^3 (1 + 3(-\tan^2 r))$, checking, since $x = -\tan^2 r$.

	n odd	n even
$A_n(0) =$	1	1
$B_n(0) =$	n	n
$C_n(0) =$	1	1
$D_n(0) =$	1	1
$A_n(1) =$	$(1-a)^{(n^2-1)/4}$	$(1-a)^{n^2/4}$
$B_n(1) =$	$(-1)^{(n-1)/2}(1-a)^{(n^2-1)/4}$	$(-1)^{n/2}(1-a)^{(n^2-4)/4}$
$C_n(1) =$	$(-1)^{(n-1)/2}(1-a)^{(n^2-1)/4}n$	$(-1)^{n/2}(1-a)^{n/4}$
$D_n(1) =$	$(1-a)^{(n^2-1)/4}$	$(1-a)^{n^2/4}$
$A_n(1/a) =$	$((1-a)/a)^{(n^2-1)/4}$	$(-1)^{n/2}((1-a)/a)^{n^2/4}$
$B_n(1/a) =$	$(-1)^{(n-1)/2}((1-a)/a)^{(n^2-1)/4}$	$((1-a)/a)^{(n^2-4)/4}n$
$C_n(1/a) =$	$((1-a)/a)^{(n^2-1)/4}$	$(-1)^{n/2}((1-a)/a)^{n^2/4}$
$D_n(1/a) =$	$(-1)^{(n-1)/2}((1-a)/a)^{(n^2-1)/4}n$	$((1-a)/a)^{(n^2-4)/4}$

We deduce the following theorems from these results.

THEOREM 16.1. *Let $\mathfrak{p}$ be a prime ideal of the second category which is not of the first category. Then $\mathfrak{p}$ never divides the sequence (A). Furthermore $\mathfrak{p}$ divides (B), (C) or (D) according as $x_0 \equiv 0$, $1 - x_0 \equiv 0$ or $1 - a_0 x_0 \equiv 0 \pmod{\mathfrak{p}}$. Its rank of apparition in every case is p, where p is the rational prime which $\mathfrak{p}$ divides.*

THEOREM 16.2. *Under the hypotheses of the preceeding theorem, if $\mathfrak{p}$ divides (C), it does not divide (D), and its rank of apparition in (B) is $2p$. Furthermore, $C_n \equiv 0 \pmod{\mathfrak{p}}$ if and only if n is an odd multiple of p. Similar results hold for divisors of (D).*

17. If we compare the results of Sections **15** and **16** for irregular prime ideals, we see that the laws of apparition are the same for all ideals save null divisors, and that the laws of apparition for (A), (C) and (D) essentially generalize Lucas' law of apparition for (V), and in a sense explain it. Furthermore, the laws of apparition for ideals of the field $\mathfrak{F}_1$ are precisely the same as for the field $\mathfrak{F}$. If in particular then x_0 and a_0 are rational integers, the four elliptic sequences are sequences of rational integers, and the ideals become ordinary primes. It is of some interest to note that the Lucas sequences associated with the primes of the first category in this case involve quadratic irrationalities, and are of the type studied in Lehmer's thesis. (D. H. Lehmer [1])

VI. Conclusion. The Laws of Repetition.

18. We conclude the paper by giving the laws of repetition for powers of primes in rational integral elliptic sequences. The extension to arbitrary algebraic integral sequences is easy, but will not be discussed here.

Assume then that x_0 and a_0 are rational integers. We need consider only regular primes p not dividing $2a_0(1-a_0)x_0(1-x_0)(1-a_0x_0)$. Let p be such a prime. Then p is odd. Assume that p divides (D), and that its rank in (B) is 2ρ. Then the rank of p in (D) is ρ, and p does not divide (A) or (C), by the results of Chapter IV. Consequently by the duplication formula (5.4) if p^k is the highest power of p dividing $B_{2\rho}$, it is also the highest power of p dividing D_ρ. Now since (B) is essentially the elliptic divisibility sequence (h) so far as regular primes are concerned, the law of repetition of powers of p in (B) follows from the results in Ward [2] for elliptic divisibility sequences; namely, if $l \leqq k$, the rank of apparition of p^l in (B) is 2ρ, and if $l \geqq k$, the rank is $2p^{l-k}\rho$. Now p is odd, and the only terms of (D) divisible by p are odd multiples of ρ. Hence since $D_n \equiv 0 \pmod{p^l}$ if and only if $B_{2n} \equiv 0 \pmod{p^l}$, we can state the following theorem:

THEOREM. *Let x_0 and a_0 be rational integers, and p a regular prime of rank ρ in (D). Furthermore, let p^k be the highest power of p dividing D_ρ. Then the rank of apparition ρ^* of p^l in (D) is ρ or $p^{l-k}\rho$ according as $l \leqq k$ or $l \geqq k$. Furthermore $D_n \equiv 0 \bmod p^l$ if and only if n is an odd multiple of ρ^*.*

Precisely similar results hold for prime divisors of (A) or (C). For the Lucas functions, these results become Lucas' law of repetition for primes in (V).

California Institute of Technology.

BIBLIOGRAPHY

R. Fricke, 1. *Die elliptischen Funktionen und ihre Anwendungen*, vol. 1, Leipzig (1916).
2. Vol. 2, Leipzig (1922).

M. Krause, 1. *Theorie der doppelperiodischen Funktionen*, Leipzig (1895).

D. H. Lehmer, 1. "An extended theory of Lucas' functions," *Annals of Mathematics*, Ser. 2, vol. 31 (1930), pp. 419-448.

M. Ward, 1. "Memoir on elliptic divisibility sequences," *American Journal of Mathematics*, vol. 70 (1948), pp. 31-74.
2. "The law of repetition of primes in an elliptic divisibility sequence," *Duke Mathematical Journal*, vol. 15 (1948), pp. 941-946.

K. Weierstrass, 1. *Werke*, vol. 1, Berlin (1894), pp. 1-49.

HARMONIC DOUBLE SERIES.*

By Leonard Tornheim.

The value of the sum c_n of the harmonic series in one variable $c_n = \sum_{p=1}^{\infty} p^{-n}$ is not known for n odd. For n even, $c_{2m} = 2^{2m-1}\pi^{2m}B_m/m!$, where B_m is the m-th Bernoulli number; in particular $c_2 = \pi^2/6$, $c_4 = \pi^4/90$, $c_6 = \pi^6/945$. We call

$$\sum_{p,q=1}^{\infty} p^{-r}q^{-s}(p+q)^{-t}$$

a harmonic double series. Its value will be denoted by (r, s, t) and we shall show how to express it as a polynomial in the c_n with rational coefficients for certain values of r, s, t.

The following lemma will be useful.

Lemma 1. *Let $f(p)$ be a monotone decreasing function with $\lim_{p\to\infty} f(p) = c$. Suppose*

$$\Phi(p, q; f) = f(q)/p(p+q) + f(p)/q(p+q) - f(p+q)/pq.$$

Then

$$\sum_{p,q=1}^{\infty} \Phi(p, q; f) = 2\sum_{r=1}^{\infty} [f(r) - c]/r^2, \tag{1}$$

$$\sum_{\substack{p,q=1\\(p,q)=1}}^{\infty} \Phi(p, q; f) = 2[f(1) - c]. \tag{2}$$

The terms in the series (1) and (2) are non-negative. For, by the monotonicity of $f(p)$

$$\Phi(p, q; f) \geqq f(p+q)[1/p(p+q) + 1/q(p+q) - 1/pq] = 0.$$

Let $g(x) = f(x) - c$ if $c \neq -\infty$. (The proof of the lemma for the case $c = -\infty$ is almost obvious and will not be discussed.) Then $\Phi(p, q; g) = \Phi(p, q; f)$, and the proof may be restricted to the case $c = 0$.

Since $1/p - 1/(p+q) = q/(p+q)p$,

---

* Received February 9, 1949.

$$\sum_{p+q<N} f(q)/p(p+q) = \sum_{q<N-1} (f(q)/q) \sum_{p<N-q} (1/p - 1/(p+q))$$
$$= \sum_{q<N-1} (f(q)/q) \sum_{s=1}^{q} (1/s - 1/(N-s)).$$

The last equality is obtained by noticing that

$$\sum_{s=1}^{N-q-1} 1/s + \sum_{s=N-q}^{N-1} 1/s = \sum_{s=1}^{q} 1/s + \sum_{s=q+1}^{N-1} 1/s;$$

thus

$$\sum_{p=1}^{N-q-1} (1/p - 1/(p+q)) = \sum_{s=1}^{N-q-1} 1/s - \sum_{s=q+1}^{N-1} 1/s$$
$$= \sum_{s=1}^{q} 1/s - \sum_{s=N-q}^{N-1} 1/s = \sum_{s=1}^{q} (1/s - 1/(N-s)).$$

Next

$$\sum_{p+q<N} f(p+q)/pq = \sum_{r=2}^{N-1} \sum_{q=1}^{r-1} f(r)/(r-q)q$$
$$= \sum_{r} (f(r)/r) \sum_{q} (1/q + 1/(r-q))$$
$$= \sum_{r=2}^{N-1} (f(r)/r)\, 2 \sum_{q=1}^{r-1} 1/q;$$

thus

$$\sum_{p+q<N} \Phi(p,q;f) = 2 \sum_{q<N-1} (f(q)/q) \sum_{s=1}^{q} (1/s - 1/(N-s))$$
$$- 2 \sum_{2 \leqq q < N} (f(q)/q) \sum_{s=1}^{q-1} 1/s$$
$$= 2 \sum_{q<N-1} (f(q)/q)(1/q) - 2f(N-1)/(N-1) \sum_{s=1}^{N-2} 1/s$$
$$- 2 \sum_{q<N-1} (f(q)/q) \sum_{s=1}^{q} 1/(N-s).$$

It remains to be shown that the last two terms approach 0 with increasing N. The absolute value of the middle term is at most

$$(2f(N-1)/(N-1))\{1 + \int_{1}^{N-1} (1/t)\,dt\} \to 0.$$

Half of the last term can be decomposed into two summations,

$$-\sum_{q<N-1} = -\sum_{q<q_0} - \sum_{q=q_0}^{N-2}$$

of the summands $(f(q)/q) \sum_{s=1}^{q} (1/(N-s))$. Choose q_0 large enough so that $f(q_0) < \epsilon$, and take N greater than $2q_0^2$. Then

$$\sum_{q<q_0} \leqq f(1) \sum_{q<q_0} 1/q \sum_{s=1}^{q} 1/(N/2) \leqq f(1)/q_0.$$

Finally

$$\begin{aligned} \sum_{q=q_0}^{N-2} &\leqq \sum_{q} (f(q_0)/q) \int_1^{q+1} ds/(N-s) \\ &= -f(q_0) \sum 1/q \log (N-1-q)/(N-1) \\ \leqq f(q_0) &\int_{q_0}^{N-1} \{1/(N-1) + q/2(N-1)^2 + q^2/3(N-1)^3 + \cdots\} dq \\ \leqq c_2 f(q_0). \end{aligned}$$

This completes the proof of the lemma.

Equation (2) may be proved similarly.[1] Also (1) can be derived from (2) by replacing p by λp, q by λq, and taking $\sum_{\lambda=1}^{\infty}$.

Every function $h(p)$ defined for positive integral p is a sum of two monotone functions h_i and h_d where

$$h_i(p) = \tfrac{1}{2}h(1) + \sum_{s=2}^{p} \max (0, h(s) - h(s-1)),$$

$$h_d(p) = \tfrac{1}{2}h(1) + \sum_{s=2}^{p} \min (0, h(s) - h(s-1)).$$

If either h_i or h_d is bounded then Lemma 1 is true for h replacing f.

In (1) take $f(p) = p^{-1}$ to obtain

COROLLARY 1. $(1, 1, 1) = 2c_3$.

Also

$$\sum_{(p,q)=1} 1/[pq(p+q)] = 2.$$

In fact, $\sum_{\substack{p,q=1 \\ (p,q)=1}}^{\infty} p^{-r}q^{-s}(p+q)^{-t} = (r, s, t)/c_{r+s+t}$.

Setting $f(p) = p^{-n+2}$ in (1) yields the following corollary after using Lemma 2 below to assure the absolute convergence of the series. This property is needed to allow the rearrangement of the terms.

COROLLARY 2. *If* $n > 2$,

$$2(n-2, 1, 1) - (1, 1, n-2) = 2c_n. \tag{3}$$

We write $[r, s, t]$ for $p^{-r} q^{-s} (p+q)^{-t}$; then $(r, s, t) = \sum_{p,q=1}^{\infty} [r, s, t]$.

---

[1] This result was originally obtained geometrically by considering the areas of the triangles described by means of a diagram of Ford, "Fractions," *American Mathematical Monthly*, vol. 45 (1938), pp. 586-601. The triangles have vertices at the points $(p/q, f(q))$ where $0 \leqq p/q \leqq 1$ and $(p, q) = 1$, two vertices belonging to the same triangle if $|p_1q_2 - p_2q_1| = 1$. The proof given here was found jointly with N. J. Fine.

The solutions to be obtained will be based only on (3) and the following simple relationships

$$(r, s-1, t+1) + (r-1, s, t+1) = (r, s, t) \tag{4}$$

resulting from the fact that the same expression in brackets is an identity;

$$(r, s, t) = (s, r, t); \tag{5}$$

$$(r, s, 0) = c_r c_s; \tag{6}$$

and

$$(r, 0, s) + (s, 0, r) = c_r c_s - c_{r+s}, \tag{7}$$

which comes from the fact that

$$\begin{aligned}\Sigma p^{-r}(p+q)^{-s} + \Sigma p^{-s}(p+q)^{-r} &= \sum_{u>p} p^{-r}u^{-s} + \sum_{u<p} u^{-s}p^{-r} \\ &= \sum_{p,u=1}^{\infty} p^{-r}u^{-s} - \sum_{u=1}^{\infty} u^{-r-s}.\end{aligned}$$

LEMMA 2. *A necessary and sufficient condition that (r, s, t) be finite is that $r+t>1$, $s+t>1$, and $r+s+t>2$.*

Proof. Observe that

$$\begin{aligned}(r, s, t) &> \sum_{q<p} \{(p+q)/k_r\}^{-r} q^{-s}(p+q)^{-t} \\ &> k_r^{\,r} \sum_{p=2}^{\infty} \sum_{q=1}^{p-1} q^{-s}(k_{-r-t}\,p)^{-r-t} \left(k_i = \begin{matrix} 1 & (i \geqq 0) \\ 2 & (i < 0) \end{matrix} \right)\end{aligned}$$

which is ∞ if $r+t \leqq 1$. Otherwise it is greater than

$$k_r^{\,r} k_{-r-t}^{\;-r-t} \sum_{p=2}^{\infty} \int_1^{p/2} q^{-s}p^{-r-t}\,dq,$$

which is ∞ if $r+s+t \leqq 2$. On the other hand if the conditions of the lemma are satisfied,

$$\begin{aligned}(r, s, t) &< \sum_{p \geqq q} p^{-r}q^{-s}(k_t p)^{-t} + \sum_{q \geqq p} p^{-r}q^{-s}(k_t q)^{-t} \\ &< k_t^{\,-t}\Big\{c_{r+t} + c_{s+t} + \sum_{q=2}^{\infty} \int_{q-1}^{\infty} q^{-s}p^{-r-t}\,dp + \sum_{p=2}^{\infty} \int_{p-1}^{\infty} p^{-r}q^{-s-t}\,dq\Big\}\end{aligned}$$

and this last expression is finite.

We shall assume that r, s, t in (r, s, t) are all integral. If every (r, s, t) is known for r, s, t all non-negative and $r+s+t=n$, by means of (4), (5), and Lemma 2 it is possible to evaluate (r_1, s_1, t_1) for any integral r_1, s_1, t_1 with $r_1 + s_1 + t_1 = n$. Hereafter we shall consider only non-negative integral r, s, t. In view of (5) we shall usually state our results with $r \geqq s$.

THEOREM 1.

$$(n-1, 1, 0) = \infty, \qquad (n-1, 0, 1) = \infty, \qquad (n, 0, 0) = \infty.$$

These follow from Lemma 2.

THEOREM 2. *If* $n \geqq 3$,

$$(1, 1, n-2) = (n-1)c_n - \sum_{a=2}^{n-2} c_a c_{n-a}, \tag{8}$$

$$(n-2, 1, 1) = \tfrac{1}{2}\{(n+1)c_n - \sum_{a=2}^{n-2} c_a c_{n-a}\}, \tag{9}$$

$$(1, 0, n-1) = \tfrac{1}{2}\{(n-1)c_n - \sum_{a=2}^{n-2} c_a c_{n-a}\}. \tag{10}$$

First from (4), $(a, 0, n-a) + (a-1, 1, n-a) = (a, 1, n-a-1)$ so that summing on a from 2 to $n-2$,

$$\sum_{a=2}^{n-2} (a, 0, n-a) + (1, 1, n-2) = (n-2, 1, 1).$$

But from (7)

$$2\sum_{a=2}^{n-2} (a, 0, n-a) = \sum_{a=2}^{n-2} c_a c_{n-a} - (n-3)c_n,$$

so that

$$\sum_{a=2}^{n-2} c_a c_{n-a} - (n-3)c_n + 2(1, 1, n-2) = 2(n-2, 1, 1).$$

Using this result simultaneously with the equation of Corollary 2 of Lemma 1, the first two equations of the theorem are readily established. Equation (10) follows from (4), (5), and (8).

THEOREM 3.

$$(a, n-a-1, 1) = (-1)^a\{\sum_{i=2}^{a} (-1)^i c_i c_{n-i} + \tfrac{1}{2}\sum_{j=2}^{n-2} c_j c_{n-j} - \tfrac{1}{2}(n+1)c_n\}. \tag{11}$$

All that is needed to prove this are equations (9), (5), and the equation

$$(a, n-a-1, 1) + (a-1, n-a, 1) = (a, n-a, 0) = c_a c_{n-a},$$

this following directly from (4) and (6).

COROLLARY. *If n is even, i. e., $n = 2m$,*

$$\sum_{k=1}^{m-1} c_{2k}c_{n-2k} = \tfrac{1}{2}(n+1)c_n.$$

It is a consequence of the fact that $(a, n-a-1, 1) = (n-a-1, a, 1)$. This identity can be translated into one for Bernoulli numbers.

Hereafter $[m]$ means the largest integer not exceeding m.

THEOREM 4. *For n odd, i. e., $n = 2m+1$, and $a > n/2$,*

$$(a, n-a-2, 2) = (-1)^a(2a-2m+1)\{\sum_{k=1}^{[m/2]} c_{2k}c_{n-2k} - \tfrac{1}{4}(n+1)c_n\} + (-1)^a \sum_{i=m+1}^{a} (-1)^i(a-i+1)c_ic_{n-i}.$$

Proof. From

$$(j, n-j-2, 2) = (j, n-j-1, 1) - (j-1, n-j-1, 2)$$

by summing get

$$(a, n-a-2, 2) = \sum_{j=m+1}^{a} (-1)^{j-a}(j, n-j-1, 1) + (-1)^{m-a}(m, m-1, 2).$$

Use the previous theorem, and the fact that $(m, m-1, 2) = \frac{1}{2}(m, m, 1)$ by equation (4), and also that $\sum_{i=2}^{n-2} c_ic_{n-i} = 2\sum_{i=2}^{m} c_ic_{n-i}$.

THEOREM 5. *If $n > 1$, $(0, 0, n) = c_{n-1} - c_n$.*

Proof. $\sum_{p,q} 1/(p+q)^n = \sum_{r=1}^{\infty} \sum_{p+q=r} 1/r^n = \sum_{r} (r-1)/r^n = c_{n-1} - c_n.$

THEOREM 6. *For n even, i. e., $n = 2m$, $(m, 0, m) = \frac{1}{2}(c_m{}^2 - c_{2m})$.*

Proof. Immediate from (7).

Denote $(a, 0, n-a)$ by f_a. Every quantity (r, s, t) for fixed $n = r+s+t$ and r, s, t all non-negative integers may be expressed as a linear combination of the f_a $(a = 0, \cdots, n)$. For, if $r \neq 0 \neq s$, then by (4) (r, s, t) may be expressed in terms of similar expressions having larger t. This process is repeated, if necessary, until only expressions (r', s', t') are obtained with either $r' = 0$ or $s' = 0$. In the next lemma the f_a are expressed in terms of the quantities $d_i = (i, i, n-2i)$ $(i = 1, \cdots, [n/2]-1)$ and polynomials in the c_j. Hence (r, s, t) may be expressed in the same way.

Using the explicit formulas of the next lemma and the relation (7), we obtain linear equations for the d_i which are solved in certain cases to give finally Theorems 7 and 8.

LEMMA 3. *$(a, 0, n-a) = \frac{1}{2}M_a + L_a$ $(a = 2, \cdots, n-2)$, where*

$$M_a = \sum_{j=1}^{\min(a,m)} (-1)^{a-j}\{{}_jC_{a-j} + {}_{j-1}C_{a-j-1}\}d_j$$

and L_a is a polynomial in $c_2, \cdots, c_n$, and $m = [n/2]$.

Denote $[a, b, c] + [b, a, c]$ by $[a, b, c]'$. We shall prove the corresponding identity in p, q when $[a, b, c]'$ replaces $2(a, b, c)$ and L_a is a linear combination of $[j, n-j-1, 1]$ and $[j, n-j, 0]$. Then summation over all positive p, q followed by the use of (11) and (6) will yield the lemma.

The proof will be by an induction on a. For $a = 2$, we have

$$[2, 0, n-2]' = \tfrac{1}{2}\{[2, 2, n-4]' - 2[1, 1, n-2]'\},$$

the validity of which can be easily verified. Likewise the identity holds for $a = 3$:

$$[3, 0, n-3]' = \tfrac{1}{2}\{[3, 3, n-6]' - 3[2, 2, n-4]'\}.$$

Next, to make the induction step, assume the result for a and $a-1$, and prove it for $a+1$. Multiply both sides of the identities for a and $a-1$ by $[1, 1, -2]$ except for terms of the form $[r, s, 0]$ or $[r, s, 1]$, these terms being multiplied by

$$[1, -1, 0] + 2[0, 0, 0] + [-1, 1, 0] \tag{11.5}$$

which equals $[1, 1, -2]$, and then subtract. An application of (4) to the left hand side and a simple manipulation of the other side using the identity ${}_rC_s + {}_rC_{s-1} = {}_{r+1}C_s$ proves the identity in the summands $[r, s, t]$.

We must show finally that no term (r, s, t) in L_a is infinite; i.e., if $t = 1$ then $r > 0$ and $s > 0$, while if $t = 0$ then $r > 1$ and $s > 1$. But the only d_i having $t = 0$ or 1 is $d_m = (m, m, t)$ and it first occurs in the induction when $a = m$. Each subsequent multiplication by (11.5) reduces r or s by 1. Only $n - m - 2$ more steps are required to reach the last step when $a = n - 2$, and r and s then still satisfy the required conditions. The proof of the lemma is complete.

When the equations of the lemma are combined with (7), $m-1$ equations in the $m-2$ unknowns d_j are obtained. Write $N_a = M_a + M_{n-a}$; then N_a is known in terms of $c_2, \cdots, c_n$ by Lemma 3 and equation (7). A relationship of dependence is easily noticed by adding the N_a; from Lemma 3 the sum is simply zero.

Ultimately we shall have to work with N_a since its value in terms of the c_j's is known. But it is easier to treat M_a and M_{n-a} separately and then add later.

We shall next show that for $a \leqq m$,

$$(12) \qquad d_a = 2 \sum_{i=1}^{a} {}_{2a-i-1}C_{a-1}\,(i, 0, n-i) \qquad (a = 2, \cdots, m).$$

This is in fact true for the corresponding summands

$$(13) \qquad [a, a, n-2a] = \sum_{i=1}^{a} {}_{2a-i-1}C_{a-1}\{[i, 0, n-i] + [0, i, n-i]\}.$$

We proceed by induction, equation (13) being true if $a = 1$. Multiply both sides of (13) by $[1, 1, -2]$. On the right are obtained terms $[n+1, 1, *]$. (In the notation $(b, c, *)$ the value of $*$ is to be chosen so that $b + c + * = n$.) But since (4) is true if brackets replace the parentheses,

$$[i+1, 1, *] = \sum_{j=1}^{i+1} [j, 0, *] + [0, 1, *];$$

and similarly for $[1, i+1, *]$. Hence, assuming (13),

$$[a+1, a+1, *] = \sum_{i=1}^{a} {}_{2a-i-1}C_{a-1}\{\sum_{j=1}^{i+1} [j, 0, *]' + [0, 1, *]'\}$$
$$= \sum_{j=2}^{a+1} \sum_{i=j-1}^{a} {}_{2a-i-1}C_{a-1}[j, 0, *]' + 2 \sum_{i=1}^{a} {}_{2a-i-1}C_{a-1}[1, 0, *]'.$$

But ${}_kC_a + {}_kC_{a-1} = {}_{k+1}C_a$, so that

$$[a+1, a+1, *] = \sum_{j=2}^{a+1} {}_{2a+1-j}C_a[j, 0, *]' + (2a/a)\,{}_{2a-1}C_a[1, 0, *]'$$
$$= \sum_{j=1}^{a+1} {}_{2(a+1)-j-1}C_{(a+1)-1}[j, 0, *]',$$

as was to be proved.

We next take the corresponding sums for M_{n-i},

$$\sum_{i=1}^{a} {}_{2a-i-1}C_{a-1}M_{n-i} = \sum_{i=1}^{a} {}_{2a-i-1}C_{a-1} \sum_{j=1}^{m} (-1)^{n-i-j}\,\{{}_jC_{n-i-j} + {}_{j-1}C_{n-i-j-1}\}d_j$$

where $a \leqq m$ and prove

LEMMA 4. $\sum_{i=1}^{a} {}_{2a-i-1}C_{a-1}M_{n-i} = \sum_{j=1}^{m} b_{aj}d_j \qquad (a = 2, \cdots, m)$
where $b_{aj} = 0$ *if*

$$(14) \qquad a < n - 2j,$$

or

$$(15) \qquad j > a \text{ and } j > n - 2a.$$

The coefficient of d_j is

$$
\begin{aligned}
b_{aj} &= \sum_{i=1}^{a} {}_{2a-i-1}C_{a-1}(-1)^{n-i-j}\{{}_jC_{n-i-j} + {}_{j-1}C_{n-i-j-1}\} \\
&= \frac{(j-1)!}{(a-1)!} \sum_{i=n-2j}^{a} (-1)^{n-i-j} \frac{(2a-i-1)!(n-i)}{(n-i-j)!(2j+i+n)!(a-i)!}
\end{aligned} \tag{16}
$$

and the factor involving the summation sign becomes, on setting $k = n - i - j$,

$$
\sum_{k=k_0}^{j} (-1)^k \frac{(2a+k+j-n-1)!(k+j)}{k!(j-k)!(a+j-n+k)!}, \tag{17}
$$

where $k_0 = \max(n-a-j, 0)$. Now since $2a+k+j-n-1 \geqq a+j-n+k$ we have that

$$
\begin{aligned}
\frac{(2a+k+j-n-1)!(k+j)}{(a+j-n+k)!} \\
= (2a+k+j-n-1)(2a+k+j-n-2) \\
\cdots (a+j-n+k+1)(k+j)
\end{aligned} \tag{18}
$$

and the right hand side considered as a polynomial in k has degree a and is expressible as

$$
\sum_{l=0}^{a} A_l\, k!/(k-l)! \tag{19}
$$

where the A_l are constants, and $k!/(k-l)! = k(k-1)\cdots(k-l+1)$ if $l \geqq 1$ and 1 if $l = 0$.

Setting $k = 0, 1, 2, \cdots$ in turn we see that

$$
A_l = 0 \qquad (n+1-2a-j \leqq 0 \leqq l \leqq n-a-j-1). \tag{20}
$$

Then (17) becomes

$$
\begin{aligned}
&\sum_{k=k_0}^{j} \sum_{l=0}^{a} (-1)^k A_l/(j-k)!(k-l)! \\
&\qquad = \sum_{l=0}^{a} \frac{A_l}{(j-l)!} \sum_{k=k_l}^{j} (-1)^k \frac{(j-l)!}{(j-k)!(k-l)!},
\end{aligned} \tag{21}
$$

where $k_l = \max(n-a-j, l)$. Here we have taken $k \geqq l$ since $k!/(k-l)! = k(k-1)\cdots(k-l+1)$ and is zero if $0 \leqq k < l$ (see (19)). Thus the left hand side of (21) equals

$$
\sum_{l=0}^{a} \frac{A_l}{(j-l)!} \sum_{k=k_l}^{j} (-1)^k {}_{j-l}C_{j-k}. \tag{22}
$$

We know from equation (16) that b_{aj} is zero if (14) holds. We want to get the conditions of (15).

By (20), $A_l = 0$ for $0 \leqq l \leqq n - a - j - 1$ if $n + 1 - 2a - j \leqq 0$ or

(23) $$j > n - 2a.$$

Also

$$\sum_{k=l}^{j} (-1)^k {}_{j-l}C_{j-k} \qquad (l \geqq n - a - j)$$

will equal zero unless $j = l$ since it is the binomial expansion of $(-1)^l (1-1)^{j-l}$. But $j \neq l$ if

(24) $$j > a$$

since $l \leqq a$. Thus Lemma 4 is proved.

We wish to compute the values of certain coefficients which are not zero.

LEMMA 5. *If $a = n - 2j$, then $b_{aj} = (-1)^j 2$. If $a > n - 2a$, then $b_{aa} = (-1)^a$. Finally if $j > a$ and $j = n - 2a$, then $b_{aj} = (-1)^a$.*

If in (16) we take $a = n - 2j$, we get

$$b_{n-2j,j} = {}_{a-1}C_{a-1}(-1)^{n-n+2j-j}\{{}_jC_{n-n+2j-j} + {}_{j-1}C_{j-1}\} = (-1)^j 2.$$

If (23) holds, but $j = a$ (instead of (24)), then by the reasoning preceding (24), (22) becomes $(-1)^a A_a$ since all terms are zero except when $l = a$. But by (19), A_a is the coefficient of k^a in (18), so that $A_a = 1$. Then from (16) we find $b_{aa} = (-1)^a$.

Finally if $j > a$ and instead of (23), $j = n - 2a$, then (17) becomes

$$\sum_{k=a}^{n-2a} (-1)^k (k + n - 2a)/k(n - 2a - k)!(k - a)!.$$

Denote this sum by S_n. First $S_{3a+1} = (-1)^a/a$. Next

(25) $$\begin{aligned} S_n &= \sum_{k=a}^{n-2a} (-1)^k \{1/(n - 2a - k)!(k - a)! \\ &\qquad + (n - 2a)/k(n - 2a - k)!(k - a)!\} \\ &= (-1)^a (1 - 1)^{n-3a}/(n - 3a)! \\ &\qquad + (n - 2a) \sum_k (-1)^k/k(n - 2a - k)!(k - a)!. \end{aligned}$$

Thus if $n - 3a > 0$,

$$\begin{aligned} S_n &= 0 + \sum_{k=a}^{n-2a} (-1)^k \{(n - 2a - k)/k(n - 2a - k)!(k - a)! \\ &\qquad + 1/(n - 2a - k)!(k - a)!\} \\ &= \{\sum_{k=a}^{n-1-2a} (-1)^k/k(n - 1 - 2a - k)!(k - a)!\} \\ &\qquad + (-1)^a (1 - 1)^{n-3a}/(n - 3a)! = S_{n-1}/(n - 2a - 1) \end{aligned}$$

because of the last part of (25).

We conclude that the coefficient of d_j is, when $j = n - 2a$,

$$\begin{aligned}((j-1)!/(a-1)!)S_n &= ((n-2a-1)!/(a-1)!)S_n \\ &= ((n-2a-2)!/(a-1)!)S_{n-1} \\ &= \cdots = (a!/(a-1)!)S_{3a+1} = a(-1)^a/a = (-1)^a.\end{aligned}$$

Since $(a, 0, n-a) + (n-a, 0, a)$ is known by (7), we obtain from (12) and Lemmas 3 and 4 the following simultaneous linear equations in the d_a:

$$d_a + \sum_j b_{aj} d_j = g_a, \tag{26}$$

where the b_{aj} are known constants, j ranges from $[(n-a+1)/2]$ to m, and g_a are polynomials in the c_i with rational coefficients.

THEOREM 7. *If n is odd, $(r, s, n-r-s)$ may be explicitly determined as a polynomial in $c_2, \cdots, c_n$ with rational coefficients.*

We shall treat only the case $n \equiv 1 \pmod 6$ i. e., $n = 6r + 1$. Separate the equations of (26) into two groups according as $a \leqq 2r$ or not. In the first case $b_{aj} = 0$ if $j \geqq n + 1 - 2a$ from (15), and (26) need be summed only to $n - 2a$.

Also $b_{aj} = 0$ if simultaneously

$$j \geqq n + 1 - 2a, \tag{27}$$

$$j > a. \tag{28}$$

If $a \leqq 2r$ then (27) implies (28), for $j > n - 2a \geqq 6r + 1 - 4r > 2r \geqq a$, whereas if $a > 2r$ then (27) is a consequence of (28), for $j \geqq a + 1 \geqq 2r + 2 = (6r + 1) + 1 - 2(2r) > n + 1 - 2a$.

To solve the system (26) first take $a = 2r + 1$. The summation in (26) is from $2r$ to $2r + 1$, and by Lemma 5 the left hand side of this equation reduces to $2d_{2r}$; thus $d_{2r} = \frac{1}{2}g_{2r+1}$. To find the other d_a we divide all equations (26) into three sets of equations

I: $a = 2r + 2s$ $(s = 1, 2, \cdots)$

II: $a = 2r + 2s + 1$ $(s = 1, 2, \cdots)$

III: $a = 2r - s$ $(s = 0, 1, 2, \cdots)$,

and solve one equation at a time from the sets I, II, III in turn.

I. Assume that $d_{2r-s+1}, d_{2r-s+2}, \cdots, d_{2r+2s-1}$ are known. When $a = 2r + 2s$

the summation in (26) is from $2r - s + 1$ to $2r + 2s$. Thus all quantities are known except d_{2r+2s}. By Lemma 5, $b_{2r+2s,2r+2s} = 1$ so that the coefficient of d_{2r+2s} on the left-hand side of (28) is 2 and d_{2r+2s} may be found.

II. Now $d_{2r-s+1}, \cdots, d_{2r+2s}$ have been determined. The summation in (26) for $a = 2r + 2s + 1$ is from $2r - s$ to $2r + 2s + 1 = a$. All d's that appear are known except d_a and d_{2r-s}. By Lemma 5, the coefficient of d_a is 0 while that of d_{2r-s} is $(-1)^{2r-s}$ so that d_{2r-s} may be expressed in terms of known quantities.

III. Finally assume $d_{2r-s}, \cdots, d_{2r+2s}$ are known. If $a = 2r - s$ the summation in (26) is from $2r$ to $2r + 2s + 1$. Every d_j appearing has been evaluated except $d_{2r+2s+1}$ whose coefficient is $(-1)^a$ by Lemma 5. Thus $d_{2r+2s+1}$ may also be found.

The order in which we solve the equations (26) is this. First take $s = 0$ in Case III, then $s = 1$ successively in Cases I, II, III. Repeat with $s = 2, 3, \cdots$ until d_{3r} is determined. This occurs either in Case I with $r = 2s$ or in Case III when $r = 2s + 1$. By then $d_{2r-s+1}, \cdots, d_{3r}$ have been determined. Finally $d_1, \cdots, d_{2r-s}$ are found by using (26) with $a = 1, \cdots, 2r - s$ respectively, for these equations are simply

$$d_a + \sum_{j=2r}^{3r} b_{aj} d_j = g_a.$$

The proof that the equations (26) may be solved when $n = 6r + 3$ and $6r + 5$ is similar and will not be given. When n is even, however, many relations of dependence occur among the equations (26) and it is impossible to solve for the d_j except in the following case.

THEOREM 8. *The values of* $(2r, 2r, 2r)$ *and* $(2r - 1, 2r, 2r + 1)$ *may be found explicitly as a polynomial in* $c_2, \cdots, c_{6r}$ *with rational coefficients.*

Proof. When $a = 2r$ and $n = 6r$, equation (26) becomes $d_{2r} + 2d_{2r} = g_{2r}$ by Lemmas 4 and 5. Hence $(2r, 2r, 2r) = g_{2r}/3$. Next by (4), $(2 + 1, 2r, 2r + 1) = (2r, 2r, 2r)/2$.

If $n = 6$ we may evaluate all $(r, s, n - r - s)$ according to our theorems except $(3, 1, 2)$, $(2, 0, 4)$ and $(4, 0, 2)$. These too may be found using (4) since $(3, 1, 2) = (3, 2, 1) - (2, 2, 2)$, $(2, 0, 4) = (1, 2, 3) - (1, 1, 4)$, and $(4, 0, 2) = (4, 1, 1) - (3, 1, 2)$. We remark in particular that $(2, 2, 2) = c_6/3$.

UNIVERSITY OF MICHIGAN,
ANN ARBOR, MICHIGAN.

ESSENTIAL AND NON ESSENTIAL FIXED POINTS.*

By M. K. Fort, Jr.

1. Introduction. Let (X, d) be a compact metric space having the fixed point property. We denote by X^X the set of all continuous functions on X into X. If f and g are in X^X, we define

$$\rho(f, g) = \sup[d(f(x), g(x)) \mid x \varepsilon X].$$

It is well known that (X^X, ρ) is a complete metric space.

Let p be a fixed point of $f \varepsilon X^X$. We say that p is an *essential fixed point* of f if corresponding to each neighborhood U of p there is $\epsilon > 0$ such that g has a fixed point in U whenever $g \varepsilon X^X$ and $\rho(f, g) < \epsilon$.

The concept of essentiality for fixed points is a stability property which is quite analogous to the concept of stable value of a function. (See [1], p. 74). Several definitions of stable fixed points have been given in the past. However, no definition of stable fixed point with which the author is familiar is equivalent to the definition of essential fixed point which is given here. We will later prove that there exists a close connection between essential fixed points and essential mappings in the sense of homotopy.

2. An approximation theorem. Members of X^X, all of whose fixed points are essential, are especially well-behaved functions. We shall prove that each element of X^X can be approximated arbitrarily closely by such functions.

We shall make use of the hyperspace $(2^X, H)$ and of the theory of semi-continuous functions into 2^X. The points of 2^X are the non-empty compact subsets of X. If $A \varepsilon 2^X$ and $\epsilon > 0$, we denote by $U(\epsilon, A)$ the set of all points $x \varepsilon X$ for which there exists $y \varepsilon A$ such that $d(x, y) < \epsilon$. The set 2^X is metrized by H, where

$$H(A, B) = \inf[\epsilon \mid A \subset U(\epsilon, B) \text{ and } B \subset U(\epsilon, A)]$$

whenever A and B are in 2^X. It is well-known that $(2^X, H)$ is a compact metric space.

A function T on a topological space S into 2^X is upper semi-continuous [lower semi-continuous] at $p \varepsilon S$ if corresponding to each $\varepsilon > 0$ there is a

---

* Received February 21, 1949.

neighborhood V of p such that $T(x) \subset U(\epsilon, T(p))$ $[T(p) \subset U(\epsilon, T(x))]$ for all $x \varepsilon V$. It is clear that T is continuous at p (with respect to the metric H) if and only if T is both upper and lower semi-continuous at p.

We next define a function F on X^X into 2^X by defining $F(f)$ to be the set of all fixed points of f whenever $f \varepsilon X^X$. The following lemma is due to Wehausen [2].

LEMMA 1. *The function F is upper semi-continuous.*

Proof. Let $f \epsilon X^X$ and suppose $\epsilon > 0$. Define

$$\delta = \inf[d(x, f(x)) \mid x \varepsilon X - U(\epsilon, F(f))]$$

unless $X = U(\epsilon, F(f))$, in which case define $\delta = 1$. It is easily verified that $\delta > 0$, and that if $g \varepsilon X^X$ and $\rho(f, g) < \delta$ then $F(g) \subset U(\epsilon, F(f))$. This proves that F is upper semi-continuous.

LEMMA 2. *Each fixed point of $f \varepsilon X^X$ is essential if and only if f is a point of continuity of F.*

Proof. Assume that each fixed point of f is essential. Let ϵ be any positive number. For each $x \varepsilon F(f)$ there is a neighborhood $V(x)$ of f such that if $g \varepsilon V(x)$ then g has a fixed point in the $\epsilon/2$-neighborhood of x. There exists a finite set $x_1, \cdots, x_n$ of points of $F(f)$ such that each point of $F(f)$ is within $\epsilon/2$ of some one of the x_k. If we now choose a neighborhood V of f which is contained in the intersection of $V(x_1), \cdots, V(x_n)$ we see that if $g \varepsilon V$ then $F(f) \subset U(\epsilon, F(g))$. We have therefore proved that F is lower semi-continuous at f. This fact, in view of Lemma 1, proves that F is continuous at f.

We now assume that F is continuous at f. Let $p \varepsilon F(f)$ and let U be a neighborhood of p. Choose $\epsilon > 0$ such that the ϵ-neighborhood of p is contained in U. Now choose $\delta > 0$ such that if $\rho(f, g) < \delta$ for $g \varepsilon X^X$, then $H(F(f), F(g)) < \epsilon$. Thus if $g \varepsilon X^X$ and $\rho(f, g) < \delta$, then g has a fixed point in U. This proves that p is an essential fixed point of f.

THEOREM 1. *If $f \varepsilon X^X$ and $\epsilon > 0$, then there exists $g \varepsilon X^X$ such that $\rho(f, g) < \epsilon$ and such that every fixed point of g is essential.*

Proof. It is known, (see [3]), that the points of continuity of an upper semi-continuous set-valued function form a residual set. Since F is upper semi-continuous and X^X is a complete metric space, we see that F is continuous at each point of a set dense in X^X. Our theorem now follows at once from Lemma 2.

3. Existence theorems. It is easy to see that there may exist members of X^X which have no essential fixed point. A simple example is obtained by taking X to be a Euclidean n-cell. For this special case, every point of X is a fixed point of the identity transformation, but no point is an essential fixed point.

It is of interest to know conditions under which a transformation will have essential fixed points. We prove two theorems in this section which give sufficient conditions for the existence of an essential fixed point. The second of these theorems, Theorem 3, contains in the hypothesis a rather strong restriction on the space X. The author does not know whether or not the conclusion of this theorem is true if the hypothesis is weakened by allowing X to be an arbitrary compact metric space.

THEOREM 2. *If f has a single fixed point, then this point is an essential fixed point of f.*

Proof. Suppose $\epsilon > 0$. By the upper semi-continuity of F, there exists a neighborhood V of f such that if $g \varepsilon V$ then $F(g) \subset U(\epsilon, F(f))$. Since $F(f)$ contains but a single point, this implies that $F(f) \subset U(\epsilon, F(g))$. Thus if $g \varepsilon V$, then $H(F(f), F(g)) < \epsilon$. We have proved that F is continuous at f, and hence it follows from Lemma 2 that the fixed point of f is an essential fixed point.

THEOREM 3. *In addition to our standing hypotheses concerning X, we shall assume that X is a topological n-manifold, either with or without boundary. If $f \varepsilon X^X$ and the set $F(f)$ of fixed points of f is totally disconnected, then f has at least one essential fixed point.*

Proof. Let us assume that f does not have an essential fixed point. We shall show that this implies the existence of a function $g \varepsilon X^X$ which does not have a fixed point. Since X has the fixed point property, this is a contradiction.

Each point $x \varepsilon X$ has a neighborhood $N(x)$ whose closure is homeomorphic to a convex subset $C(x)$ of Euclidean n-space. For each $x \varepsilon F(f)$ choose a neighborhood $U(x)$ with the following properties: $U(x) \subset N(x)$, the intersection of $F(f)$ and the boundary of $U(x)$ is vacuous, there exist members of X^X arbitrarily close to f which do not have fixed points in $U(x)$. Let us cover $F(f)$ with a finite set $U(x_1), \cdots, U(x_k)$ of these neighborhoods.

Now define $F_1 = F(f) \cap U(x_1)$ and define

$$F_i = F(f) \cap U(x_i) - \bigcup_{j=1}^{i-1} F_j$$

if $1 < i \leqq k$. The sets F_i are closed, pairwise disjoint sets; and we may suppose the sets $U(x_1), \cdots, U(x_k)$ chosen so that each set F_i is non-vacuous.

Choose open sets $V_1, \cdots, V_k, W_1, \cdots, W_k$ with the following properties: $F_i \subset V_i \subset \bar{V}_i \subset W_i \subset \bar{W}_i \subset U(x_i)$, $\bar{W}_i \cap \bar{W}_j$ is the empty set if $i \neq j$, $f(\bar{W}_i) \subset N(x_i)$.

Let h_i be a homeomorphism of $\overline{N(x_i)}$ onto $C(x_i)$ for each i, $1 \leqq i \leqq k$. Define $\epsilon = \inf[d(x, f(x)) \mid x \in X - \bigcup_{i=1}^{k} V_i]$. Choose $\delta > 0$ so that if $p, q \in C(x_i)$ and $\|p - q\| < \delta$, then $d(h_i^{-1}(p), h_i^{-1}(q)) < \epsilon$ for $1 \leqq i \leqq k$; choose $\eta > 0$ so that if $u, v \in \overline{N(x_i)}$ and $d(u, v) < \eta$, then $\| h_i(u) - h_i(v) \| < \delta$ for $1 \leqq i \leqq k$. Next, we may choose functions $g_i \in X^X$ for $1 \leqq i \leqq k$ such that $\rho(f, g_i) < \eta$, g_i does not have a fixed point in $\bar{V}_i$, and $g_i(\bar{W}_i) \subset N(x_i)$.

If $p \in W_i - \bar{V}_i$ we define $a_i(p) = \inf[d(p, x) \mid x \in \bar{V}_i]$ and $b_i(p) = \inf[d(p, x) \mid x \in X - W_i]$. We assume that each W_i has been chosen small enough so that $X - W_i$ is non-vacuous. Let $A_i(p) = a_i(p)/(a_i(p) + b_i(p))$ and $B_i(p) = b_i(p)/(a_i(p) + b_i(p))$. We now define $g(p)$ to be $f(p)$, if $p \in X - \bigcup_{i=1}^{k} W_i$; $g_i(p)$, if $p \in \bar{V}_i$ for some i, $1 \leqq i \leqq k$; $h_i^{-1}(A_i(p) h_i f(p) + B_i(p) h_i g_i(p))$, if $p \in W_i - \bar{V}_i$ for some i, $1 \leqq i \leqq k$. The function g belongs to X^X and it is easily seen that $d(p, g(p)) > 0$ for all $p \in X$. This is a contradiction.

4. Local study of essential fixed points. Although Theorem 3 is frequently useful in proving the existence of an essential fixed point of a function f, this theorem is of no value in deciding which fixed points are essential if f has more than one fixed point. For example, let X be the unit disk in the complex plane and let $f(z) = z^2$. The fixed points of f are 0 and 1. We know by Theorem 3 that at least one of these is essential, but we do not know which ones are essential and which ones are not essential. It is desirable to have theorems which we may use to prove that a fixed point is essential merely by examining the behavior of the function in a neighborhood of the point. The following is a simple theorem of this type.

THEOREM 4. *If $f \in X^X$, $p \in X$ and p has arbitrarily small neighborhoods V such that $\bar{V}$ has the fixed property and $f(\bar{V}) \subset V$, then p is an essential fixed point of f.*

Proof. It is clear that p is a fixed point of f.

Let U be any neighborhood of p. Choose a neighborhood V of p such that $\bar{V} \subset U$, $\bar{V}$ has the fixed point property, $f(\bar{V}) \subset V$, and $X - V$ is non-vacuous. Let us define $\epsilon = \inf[d(x,y) \mid x \,\varepsilon\, f(\bar{V}), y \,\varepsilon\, X - V]$. Since $f(\bar{V})$ and $X - V$ are disjoint, non-vacuous, compact sets, we see that ϵ is a positive number.

Now let $g \,\varepsilon\, X^X$ and assume that $\rho(g,f) < \epsilon$. Then $g(\bar{V}) \subset V \subset \bar{V}$, and since $\bar{V}$ has the fixed point property, g has a fixed point in $\bar{V}$ and hence in U. This proves that p is an essential fixed point of g.

Using the above theorem, it is now easy to show that 0 is an essential fixed point of the function $f(z) = z^2$ for z in the unit complex disk.

It is very easy to describe situations where Theorem 4 is of no use. For example, consider the function $f(z) = z/|z|^{\frac{1}{2}}$ defined for all z in the unit complex disk, $f(0)$ being defined to be 0. The behavior of this function on neighborhoods of 0 is fairly typical of the situation we may expect in general, in that the function f does not transform small neighborhoods of 0 into themselves. This suggests that one might prove stronger theorems than Theorem 4 if one had as tools fixed point theorems for mappings of sets into subsets of some containing space rather than merely into subsets of themselves. We prove one such theorem in the next section.

5. A fixed point theorem.

Let n be a positive integer and let E_n denote Euclidean n-space. The set of all points x in E_n such that $\| x \| \leqq r$ will be denoted by $K(r)$, and the set of all points x in E_n such that $\| x \| = r$ will be denoted by $S(r)$. If $x \,\varepsilon\, E_n$ and x is not the origin, we let $P(x) = x/\| x \|$.

Throughout this section we shall assume that r is a positive number and that f is a continuous function on $K(r)$ into E_n such that $f(x) \neq x$ for all $x \,\varepsilon\, S(r)$. We associate with f a mapping f^* of $S(1)$ into $S(1)$ by defining $f^*(x) = P(rx - f(rx))$ for all $x \,\varepsilon\, S(1)$.

THEOREM 5. *If f^* is an essential mapping (not homotopic to a constant mapping) then f has a fixed point.*

Proof. Suppose that f does not have a fixed point. In this case $trx \neq f(trx)$ for all $x \,\varepsilon\, S(1)$ and $0 \leqq t \leqq 1$. Thus we may define $D(x,t) = P(trx - f(trx))$ for all $x \,\varepsilon\, S(1)$ and $0 \leqq t \leqq 1$. The function D is clearly a homotopy of f^* to a constant function.

Example 1. Let f be such a mapping that $f(K(r)) \subset K(r)$. It is easily seen that f^* is homotopic to the identity mapping if one considers the homotopy defined by $D(x,t) = P(rx - tf(rx))$. The Brouwer fixed point theorem now follows from the fact that the identity mapping on $S(1)$ is essential.

Example 2. Let $n=2$ and let g be the mapping of $K(1)$ into E_2 which takes the point (u,v) into $(2u, v/2)$. The homotopy defined by $D((u,v),t) = P((tu - 2u, 2v - 3tv/2))$ deforms g^* into the mapping which takes (u,v) into $(-u,v)$. This mapping, being a square root of the identity, is essential. Thus g^* is essential.

Now there exists $\epsilon > 0$ such that $\| x - g(x) \| \geqq \epsilon$ for all $x \varepsilon S(1)$. It is easy to see that if f is any mapping of $K(1)$ such that $\| f(x) - g(x) \| < \epsilon$ for all $x \varepsilon S(1)$, then f^* is homotopic to g^* and consequently is essential. Any such mapping f, by Theorem 5, has a fixed point. Notice that f is required to be close to g only on $S(1)$ and not on all of $K(1)$.

THEOREM 6. *If f^* is homotopic to a constant, then there exists a mapping g on $K(r)$ into E_n without fixed points such that $g(rx) = f(rx)$ for all $x \varepsilon S(1)$. Moreover, if $\delta > 0$, we may choose g so that $\rho(g,I) \leqq \rho(f,I) + \delta$, where I is the identity mapping on $K(r)$.*

Proof. Let D be such a homotopy that $D(x,1) = f^*(x)$ and $D(x,0)$ is constant. Define $g(trx) = trx - [t \| rx - f(rx) \| + (1-t)\delta] D(x,t)$ for all $x \varepsilon S(1)$ and $0 \leqq t \leqq 1$. It is easily seen that g is a continuous function on $K(r)$ into E_n which has no fixed points. Also,

$$g(rx) = rx - \| rx - f(rx) \| D(x,1) = rx - \| rx - f(rx) \| f^*(x) = f(rx)$$

for all $x \varepsilon S(1)$.

Moreover,

$$\begin{aligned} \| g(trx) - trx \| &= [t \| rx - f(rx) \| + (1-t)\delta] \| D(x,t) \| \\ &\leqq \| rx - f(rx) \| + \delta \leqq \rho(f,I) + \delta \end{aligned}$$

for all $x \varepsilon S(1)$ and $0 \leqq t \leqq 1$. Thus $\rho(g,I) \leqq \rho(f,I) + \delta$.

6. Isolated fixed points on manifolds.

Let X be a compact n-manifold (either with or without boundary) which has the fixed point property. Let $f \varepsilon X^X$ and let p be an isolated fixed point of f. In case X has a non-vacuous boundary, we shall also assume that p is not a boundary point of X. Let N be a neighborhood of p whose closure is homeomorphic to the closed unit sphere $K(1)$ in E_n. We may assume that h is a homeomorphism of $\bar{N}$ onto $K(1)$ which takes p into the origin of E_n. We also assume that N has been chosen small enough so that the only fixed point of f in $\bar{N}$ is p.

For all sufficiently small positive numbers r, say $0 < r \leqq R$, $fh^{-1}(K(r)) \subset N$ and hence $hfh^{-1} \mid K(r)$ is defined and is a function which we shall denote by f_r. Since the only fixed point of f_r is the origin, we may associate with each f_r, $0 < r \leqq R$, a mapping $f^*{}_r$ of $S(1)$ into $S(1)$ by defining $f^*{}_r(x) = P(rx - f_r(rx))$ for each $x \varepsilon S(1)$. It is clear that $f^*{}_r$ is homotopic to $f^*{}_R$ if $0 < r \leqq R$.

THEOREM 7. *The point p is an essential fixed point of f if and only if $f^*{}_R$ is an essential mapping.*

Proof. Let us assume that $f^*{}_R$ is an essential mapping. Let U be any neighborhood of p. Choose r, $0 < r < R$, such that $h^{-1}(K(r))$ and $fh^{-1}(K(r))$ are both contained in U. Now choose $\epsilon > 0$ such that if $\rho(f, g) < \epsilon$ then $gh^{-1}(K(r))$ is contained in both U and N. For all g such that $\rho(f, g) < \epsilon$, we may define g_r and $g^*{}_r$ in the obvious manner. We may assume that ϵ has been chosen small enough so that $g^*{}_r$ and $f^*{}_r$ are homotopic. Thus, if $\rho(g, f) < \epsilon$, $g^*{}_r$ is essential and it follows from Theorem 5 that g_r has a fixed point q in $K(r)$. The point $h^{-1}(q)$ is in U and is a fixed point of g. This proves that p is an essential fixed point of f.

Now let us assume that $f^*{}_R$ is non-essential. Let ϵ be any positive number. We will show that there exists $g \varepsilon X^X$ such that $\rho(f, g) < \epsilon$ and such that g does not have a fixed point in N. Choose r, $0 < r < R$, such that $h^{-1}(K(r))$ and $fh^{-1}(K(r))$ are both contained in the $\epsilon/4$ neighborhood of p. Now choose s, $0 < s < r$, such that $\rho(f_s, I) + 2s < r$. By Theorem 6 there exists a function g_s on $K(s)$ into E_n such that $g_s(x) = f_s(x)$ for all $x = S(s)$ and such that g_s has no fixed points. Moreover, we may assume that $\rho(g_s, I) + s \leqq \rho(f_s, I) + 2s < r$, and hence $g_s(K(s)) \subset K(r)$. We now define $g(x) = f(x)$ if $x \varepsilon X - h^{-1}(K(s))$ and $g(x) = h^{-1}g_s h(x)$ if $x \varepsilon h^{-1}(K(s))$. It is easy to verify that g has the required properties. We have thus shown that p is not an essential fixed point of f.

Example 3. Let X be the set of all points (u, v) in E_2 such that $\max(|u|, |v|) \leqq 1$. Let f be the function on X which transforms the point $(u^{1/3}, v^3)$. Locally, f behaves similarly to the mapping defined in Example 2. Combining the methods used in Example 2 with Theorem 7, it is easy to prove that the origin is an essential fixed point of f.

7. A fixed point theorem for product spaces. Let X be a compact metric space having the fixed point property, and let T be the closed unit interval of real numbers. We denote by f a function on $X \times T$ into $X \times T$, and denote the transform of the point (x, t) under f by $(f_t(x), g(x, t))$.

THEOREM 8. *If for each t, $0 \leqq t \leqq 1$, the function f_t of X into X has only essential fixed points, then f has a fixed point.*

Proof. For each t let $Z(t) = F(f_t)$. Using Lemma 2 it is easy to prove that Z is a continuous function on T into 2^X. It follows that there exists a connected set C contained in $X \times T$ such that if $(x, t) \varepsilon C$ then $x \varepsilon Z(t)$, and such that there exist $x_0, x_1 \varepsilon X$ for which $(x_0, 0) \varepsilon C$ and $(x_1, 1) \varepsilon C$. Consider the continuous real valued function on C defined by $h(x, t) = g(x, t) - t$. Since $h(x_0, 0) \geqq 0$ and $h(x_1, 1) \leqq 0$, it follows that $h(x^*, t^*) = 0$ for some $(x^*, t^*) \varepsilon C$. Thus $g(x^*, t^*) = t^*$ and $f_{t^*}(x^*) = x^*$, and we see that (x^*, t^*) is a fixed point of f.

The hypothesis of the above theorem is quite restrictive. It is not known whether or not there exist spaces X and Y with the fixed point property for which $X \times Y$ does not have the fixed point property. Consequently, one would hope for a much stronger theorem than Theorem 8.

UNIVERSITY OF ILLINOIS.

BIBLIOGRAPHY.

1. W. Hurewicz and H. Hallman, *Dimension theory*, Princeton, 1941.

2. J. V. Wehausen, "Transformations in metric spaces and ordinary differential equations," *Bulletin of the American Mathematical Society*, vol. 51 (1945), pp. 113-119.

3. M. K. Fort, Jr., "A unified theory of semi-continuity," *Duke Mathematical Journal*, vol. 16 (1949), pp. 237-246.

ON THE COMMUTATOR GROUP OF A SIMPLE ALGEBRA.*[1]

By Shianghaw Wang.

Introduction. Let A be an algebra. If α in A is a commutator: $\alpha = \beta\gamma\beta^{-1}\gamma^{-1}$, clearly, its norm (in any sense) must be unity. It follows that any element in the commutator group must be of norm 1.

Now, consider as an example the algebra D of quaternions over the field R of real numbers. Any element α in D is contained in a quadratic extension of R, which is by necessity a field $R(i)$, isomorphic to the field of complex numbers. If α is of reduced norm 1: $\alpha\bar{\alpha} = 1$, it can be written in the form $\cos\theta + i\sin\theta = e^{\theta i}$. Since there exists an element j in D such that $j\xi j^{-1} = \bar{\xi}$ for any ξ in $R(i)$, we have $e^{\theta i} = e^{\frac{1}{2}\theta i} j (e^{\frac{1}{2}\theta i})^{-1} j^{-1}$. This shows any element of reduced norm 1 in D is a commutator.

Nakayama and Matsushima[2] proved that any element of reduced norm 1 in a p-adic division algebra is a product of at most three commutators. Combining this result with the above simple fact in the quaternion algebra, we may say that any element of reduced norm 1 in a "local" division algebra is in the commutator group. It is but a step to infer the corresponding statement for any "local" simple algebra.

Our main purpose here is to prove that the statement holds "in the large," i. e., the commutator group of the group of all regular elements in a simple algebra A over an algebraic number field k coincides with the group of all elements of reduced norm 1. From this theorem and the norm theorem of Hasse-Schilling-Maass, we infer that the factor commutator group of A is isomorphic in a natural way to the group of all elements in k, which are positive at all infinite primes where A is ramified.

Whether the theorem is true for any simple algebra we cannot yet decide. Nevertheless, we shall prove its general validity in the special case where the index of the algebra is square free.

The commutator group of a simple ring which is not a pure division ring is a generalization of the classical special linear group. Apart from

---

* Received February 21, 1949.

[1] This investigation of the multiplicative group of a simple algebra was suggested to the author by Professor E. Artin.

[2] T. Nakayama and Y. Matsushima, "Über die multiplikative Gruppe einer p-adischen Divisionsalgebra," *Proceedings of the Imperial Academy of Japan*, 19, 1943.

certain special cases to be discussed in I below, the question of its structure was settled by Dieudonné.[3] But the group of a pure division ring is entirely different. The structure problem in this case seems very difficult even when a division algebra over an algebraic number field is concerned.

I. Commutator Groups of Simple Rings.

Let A be a simple ring, i. e., a full matrix ring over a division ring D. The set of all regular elements in A forms a group which will be denoted by the same symbol denoting the ring itself.

Suppose A is of degree $m > 1$ over $D: A = M_{(m)} \times D$, and let E_{ij}, i, $j = 1, \cdots, m$, form the usual matrix basis. E being the unit matrix, let $B_{ij}(\lambda) = E + \lambda E_{ij}$, $i \neq j$, and let $E(\alpha) = E + (\alpha - 1) E_{mm}$. Then the set of all $B_{ij}(\lambda)$ generates the commutator group A' of A and any coset of A' in A can be represented by a matrix $E(\alpha)$. The central elements of A or of A' are contained in the center k of D.

Dieudonné proved in [2] that the factor commutator group A/A' is isomorphic to D/D' and the isomorphism can be brought up by the function Δ which maps the coset of $E(\alpha)$ to the coset of α in D/D'. $\Delta(X)$ was called the determinant of the matrix X. The following lemma is an immediate consequence of this result of Dieudonné:

LEMMA 1. *The matrix* $\alpha_1 E_1 + \cdots + \alpha_m E_{mm}$ *is in* A' *if and only if the product* $\alpha_1 \cdots \alpha_m$ *is in* D'.

We shall call a group simple if any proper normal subgroup is contained in the center. Dieudonné determined the structure of A' by showing that it is simple when $m > 2$, or when $m = 2$ and the center k of D is not the prime field R_2, R_3, or R_5 of characteristic 2, 3, or 5 respectively. We wish to show that this restriction can be removed. In fact, we shall prove that *the group* A' *is simple, when* $m > 2$, *or when* $m = 2$ *and* $D \neq R_2$ *or* R_3.

Proof. The general proof given below does not apply when $D = R_5$. For this singular case, see Dickson [3], p. 85.[4]

Let $A = M_{(2)} \times D$, $D \neq R_2, R_3, R_5$. Let N be a normal subgroup of A, containing a matrix

$$X = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \notin A' \cap k.$$

---

[3] J. Dieudonné, "Les déterminants sur un corps non commutatif," *Bulletin de la Société Mathématique de France*, vol. 71 (1943).

[4] L. E. Dickson, *Linear Groups*, Leipzig, 1901.

Following the method in [2], we shall first construct a matrix of the form $B_{12}(\lambda_0)$ from X and its transforms.

Suppose one of β and γ is not 0. We may assume $\beta \neq 0$. For, if $\gamma \neq 0$, X may be replaced by

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^{-1} = \begin{pmatrix} \delta & -\gamma \\ -\beta & \alpha \end{pmatrix}.$$

Furthermore, we may assume $\alpha = 0$, because

$$(1) \qquad \begin{pmatrix} 1 & 0 \\ -\beta^{-1}\alpha & 1 \end{pmatrix}^{-1} \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\beta^{-1}\alpha & 1 \end{pmatrix} = \begin{pmatrix} 0 & \beta \\ \gamma_1 & \delta_1 \end{pmatrix}.$$

For $\xi \varepsilon D$, k_ξ will denote a subfield of D, containing the field $k(\xi)$, and containing at least one element outside k in case $k = R_2$, R_3 or R_5.

If $\eta \varepsilon k_\gamma$, we have

$$(2) \qquad \begin{pmatrix} 0 & \beta \\ \gamma & \delta \end{pmatrix}^{-1} \begin{pmatrix} \eta & 0 \\ 0 & \eta^{-1} \end{pmatrix}^{-1} \begin{pmatrix} 0 & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} \eta & 0 \\ 0 & \eta^{-1} \end{pmatrix} = \begin{pmatrix} \eta^2 & * \\ 0 & \beta^{-1}\eta^{-1}\beta\eta^{-1} \end{pmatrix} = Y.$$

When $k \neq R_2$, R_3, R_5, we choose η in k such that $\eta^4 \neq 1$. When $k = R_2$, R_3 or R_5, we choose η in k_γ outside k such that $\eta^2 \notin k$. In both cases, Y has the property that either its two diagonal elements are distinct or they do not belong to k.

If $\beta = \gamma = 0$ in the original matrix X, X has already the above property, because it is not in the center of A' by assumption. Hence, N always contains a matrix $\begin{pmatrix} \alpha & \beta \\ 0 & \delta \end{pmatrix}$ where $\alpha \neq \delta$ or $\alpha = \delta$ but $\alpha \notin k$. So, there exists an element $\zeta \varepsilon D$ for which $\alpha\zeta \neq \zeta\delta$. We have

$$(3) \qquad \begin{pmatrix} 1 & \zeta \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha & \beta \\ 0 & \delta \end{pmatrix} \begin{pmatrix} 1 & \zeta \\ 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \alpha & \beta \\ 0 & \delta \end{pmatrix}^{-1} = \begin{pmatrix} 1 & \zeta - \alpha\zeta\delta^{-1} \\ 0 & 1 \end{pmatrix} = B_{12}(\lambda_0),$$

where $\lambda_0 = \zeta - \alpha\zeta\delta^{-1} \neq 0$.

Now, consider a field k_{λ_0}. Starting with $B_{12}(\lambda_0)$ as the original matrix X, we first obtain a matrix of the form $\begin{pmatrix} 0 & \beta \\ \gamma & \delta \end{pmatrix}$ as in (1), $\beta, \gamma, \delta \varepsilon k_{\lambda_0}$. This time, we choose $\eta \varepsilon k_{\lambda_0}$ such that $\eta^4 \neq 1$ and obtain a matrix of the form $\begin{pmatrix} \alpha & \beta \\ 0 & \delta \end{pmatrix}$ as in (2), $\alpha, \beta, \delta \varepsilon k_{\lambda_0}$, $\alpha \neq \delta$. Taking $\zeta = (1 - \alpha\delta^{-1})^{-1}$ in (3), we obtain the matrix $B_{12}(1)$.

Let $\lambda \varepsilon D$ be arbitrary. Starting with $B_{12}(1)$ as our matrix X, we repeat the above process, keeping the coefficients of the matrices inside a field k_λ. Setting $\zeta = \lambda(1 - \alpha\delta^{-1})^{-1}$ in (3), we obtain $B_{12}(\lambda)$. Since

$$\begin{pmatrix}1 & 0\\ \lambda & 1\end{pmatrix}=\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}\begin{pmatrix}1 & \lambda\\ 0 & 1\end{pmatrix}^{-1}\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}^{-1},$$

we have finally the matrix $B_{21}(\lambda)$. So, $N=A'$ and the statement is proved.

Since, as we shall prove, the commutator group of a simple algebra over an algebraic number field coincides with the group of all elements of norm 1, it is conceivable that the group D/D' might always be isomorphic to a subgroup of the center k and Dieudonné's determinant of a matrix could then be realized by an element in k. As the following example shows, this is not the case:

Let K be the field of all power series over R_2 of the form

$$\sum_{\nu=r}^{\infty} a_\nu x^\nu, \qquad a_\nu = 0 \text{ or } 1, \qquad r \gtreqless 0.$$

We introduce a second variable y and consider all the double power series

$$\sum_{\mu=s}^{\infty} \alpha_\mu y^\mu, \qquad \alpha_\mu \varepsilon K, \qquad s \gtreqless 0.$$

We add and multiply them formally applying the rule $y\alpha = \alpha^\sigma y$, $\alpha \varepsilon K$, where σ is the automorphism of K defined by the equation

$$x^\sigma = x + x^2. \tag{4}$$

As can be readily verified, these double power series form a division ring D. Now, in a power series field, an automorphism of infinite period, which can be defined by an equation like (4), leaves fixed only the coefficient field. Since σ is obviously of infinite period, it can be easily seen from the above remark that the center of D is exactly $R_2 = \{0, 1\}$.

Let $\xi = \sum_{\mu=s}^{\infty} \alpha_\mu y^\mu$ be in D, $\alpha_s \neq 0$. We define $|\xi| = y^s$. By the rule of multiplication, we have $|\xi\eta| = |\xi|\,|\eta|$. So, this is a homomorphism of D onto the infinite cyclic group generated by y. Since the latter is abelian, D/D' is certainly not trivial.

II. Proof of the Main Theorem.

Let A be a central simple algebra over k. By the norm $N\alpha$ of $\alpha \varepsilon A$ we mean the reduced norm $N_{Ak}(\alpha)$ of α over k. If α is in the maximal subfield K of A, then $N\alpha = N_{Kk}(\alpha)$.

Let K be a cyclic maximal subfield over k and let σ be a generating automorphism. Then there exists an element τ in A such that $\tau\xi\tau^{-1} = \xi^{\sigma}$ for any $\xi \varepsilon K$. If $N\alpha = 1$, $\alpha \varepsilon K$, there exists an element ξ in K such that $\alpha = \xi^{\sigma-1}$ by a theorem of Hilbert. So, $\alpha = \tau\xi\tau^{-1}\xi^{-1}$ and we have proved

LEMMA 2. *Any element of norm* 1 *in* A, *that is contained in a cyclic maximal subfield, is a commutator.*

Let D be a central division algebra over k and F a field of degree m over k.

LEMMA 3. *If* $\alpha \varepsilon D$ *is in the commutator group* D'_F *of* D_F, *then* α^m *is in* D'.

Proof. Consider the algebra $A = M_{(m)} \times D$. Since F is isomorphic to maximal subfield of $M_{(m)}$, we may regard $D_F = F \times D$ as a subalgebra of A. Hence, $\alpha \varepsilon D'_F \subset A'$. By Lemma 1, $\alpha^m \varepsilon D'$.

Now, let K be a splitting field of $D = D_{(n)}$, of degree n over k. If $\alpha \varepsilon D$ has norm 1 in D, it still has norm 1 in $D_K = M_{(n)}$. But an element of norm 1 in a full matrix algebra is nothing but a matrix of determinant 1. So, $\alpha \varepsilon D'_K$. This, together with Lemma 3, proves

LEMMA 4. *If* $N\alpha = 1$ *in* $D = D_{(n)}$, *then* $\alpha^n \varepsilon D'$.

Let k be a p-adic field and $D = D_{(n)}$ be a central division algebra over k. Then D contains an unramified maximal subfield W. For a suitably chosen generating automorphism σ of W and a suitably chosen prime element π in D, we have

$$D = W + W\pi + \cdots + W\pi^{n-1}; \; \pi\xi\pi^{-1} = \xi^{\sigma} \text{ for } \xi \varepsilon W, \; \pi^n = p.$$

So, D admits an irreducible representation over W, in which the element $\alpha = \alpha_0 + \alpha_1\pi + \alpha_2\pi^2 + \cdots + \alpha_{n-1}\pi^{n-1}$, $\alpha_i \varepsilon W$, is represented by the matrix

$$\begin{pmatrix} \alpha_0 & \alpha_1 & \cdots & \alpha_{n-1} \\ p\alpha^{\sigma}{}_{n-1} & \alpha^{\sigma}{}_0 & \cdots & \alpha^{\sigma}{}_{n-2} \\ p\alpha_1{}^{\sigma^{n-1}} & p\alpha_2{}^{\sigma^{n-1}} & \cdots & \alpha_0{}^{\sigma^{n-1}} \end{pmatrix}.$$

It can be shown that α is integral if and only if so are $\alpha_0, \alpha_1, \cdots, \alpha_{n-1}$.

Now, let α be a unit in D. Let q be the number of residue-classes of p in k. We have $\alpha_0 \equiv \omega \pmod{p}$, ω being a $(q^n - 1)$-th root of unity in W. It follows from the above representation that $N\alpha \equiv N\alpha_0 \equiv N\omega \pmod{p}$.

The following lemma is a slight improvement of a result of Nakayama and Matsushima, referred to in the introduction:

LEMMA 5. *Any element α of norm 1 in D is a product of at most two commutators.*

Proof. Since α is of norm 1, it is a unit. Using the notations introduced above, we have $\alpha \equiv \omega \pmod{\pi}$ and $1 = N\alpha \equiv N\omega \pmod{p}$. Since $N\omega$ is a $(q^n - 1)$-th root of unity, the latter congruence implies $N\omega = 1$. So, $\omega^{1+q+\cdots+q^{n-1}} = 1$, or ω is a $(q^n - 1)/(q - 1)$-th root of unity.

Suppose ω is a *primitive* $(q^n - 1)/(q - 1)$-th root of unity. If $\nu < n$, we have $q^\nu - 1 < (q^n - 1)/(q - 1)$. So, ω is not a $(q^\nu - 1)$-th root of unity for $\nu < n$, and therefore generates the residue-class field of π in D (i. e., the residue-class field of p in k by adjoining the class of ω). It follows that α also generates the residue-class field. Consequently, α generates a maximal unramified subfield of D. Since α is in a cyclic maximal subfield, it is a commutator.

If ω is not primitive, let $\omega_1 \varepsilon W$ be a primitive $(q^n - 1)/(q - 1)$-th root of unity. Then $\omega_1^{-1}\alpha \equiv \omega_1^{-1}\omega \pmod{\pi}$, and $\omega_1^{-1}\omega$ is primitive. Since both ω_1 and $\omega_1^{-1}\alpha$ are commutators, α is in this case a product of two commutators.

Observe that, if we have proved that any element of norm 1 in D is in D', D being an arbitrary division algebra, we may conclude immediately that any element of norm 1 in $A = M_{(m)} \times D$ is in A'. For, any element $\alpha \varepsilon A$ is equivalent to a diagonal matrix $E(\beta)$ mod A'. If $N\alpha = 1$, we have $N_{Dk}(\beta) = 1$ and $\beta \varepsilon D'$. Consequently, $E(\beta) \varepsilon A'$ and $\alpha \varepsilon A'$.

Let $A = A_{(n)}$ be a central simple algebra over the p-adic field k, and $\omega_1, \omega_2, \cdots, \omega_{n^2}$ form a basis. Let $\xi = t_1\omega_1 + \cdots + t_{n^2}\omega_{n^2}$ be a general element of A. We valuate A by defining $|\xi| = \max(|t_1|, \cdots, |t_{n^2}|)$. This is a valuation of A in a modified sense, inasmuch as the condition $|\xi\eta| = |\xi|\,|\eta|$ for an orthodox valuation is not necessarily satisfied. Nevertheless, it turns A into a complete normed vector space over k, and different bases give rise to topological equivalent spaces. Needless to say, any subalgebra is a linear subspace and therefore, as a point set, is closed in A. Furthermore, if the subalgebra is a field, the valuation introduced above induces in it the same topology as its standard valuation does.

As in any normed vector space, $\xi + \eta$, $\xi - \eta$, $a\xi$ are of course continuous functions of ξ, η. Since the coefficients of $\xi\eta$ are polynomials in those of ξ, η, $\xi\eta$ is also a continuous function of ξ, η. Let $f_\xi(x) = x^n + a_1(t)x^{n-1} + \cdots + a_n(t)$ be the principal polynomial of ξ. Then $a_1(t), \cdots, a_n(t)$, as polynomials of $t_1, \cdots, t_{n^2}$, are continuous functions of ξ. If ξ is regular, then $a_n(t) \neq 0$ and ξ^{-1} may be written in the form $\xi^{-1} = -a_n(t)^{-1}(\xi^{n-1} + a_1(t)\xi^{n-2} + \cdots + a_{n-1}(t)$. Consequently, ξ^{-1} is a continuous function of ξ.

LEMMA 6. *If α generates a maximal subfield K of A, then all elements in a sufficiently small neighborhood of α generate maximal subfields which are isomorphic to K.*

Proof. Since there are only finitely many intermediate fields between K and k, and they are closed in K, there exists a neighborhood U_K of α in K, which does not intersect them. Since $f_\alpha(x)$ is irreducible over k, it has no multiple roots. So, there exists a neighborhood U_A of α in A such that, if $\xi \varepsilon U_A$, $f_\xi(x)$ will be sufficiently near $f_\alpha(x)$ so that it will have a root θ in U_K on account of Hensel's Lemma. Since θ does not belong to the intermediate fields, it generates K: $K = k(\theta)$. But then $f_\xi(x)$ must be irreducible. Consequently, $k(\xi)$ is a maximal subfield and is isomorphic to $k(\theta) = K$.

Let A be a central simple algebra over the algebraic number field k. Then $A^p = A \times k_p$ is the p-adic completion of A, p being a prime in k. If $\alpha_i \varepsilon A^{p_i}$, $i = 1, \cdots, t$, are given, we can approximate them simultaneously by an element ξ in A. In fact, we need only approximate the coefficients.

LEMMA 7. *Let $\epsilon > 0$ be given and let $\alpha_i \varepsilon A^{p_i}$, $i = 1, \cdots, t$, be elements of norm 1. There exists an element $\xi \varepsilon A'$ such that $|\xi - \alpha_1|^{p_1} < \epsilon, \cdots,$ $|\xi - \alpha_t|^{p_t} < \epsilon$.*

Proof. Since α_i is of norm 1, it is in $(A^{p_i})'$. Let

$$\alpha_i = \beta_{i1}\gamma_{i1}\beta_{i1}^{-1}\gamma_{i1}^{-1} \cdots \beta_{ix_i}\gamma_{ix_i}\beta_{ix_i}^{-1}\gamma_{ix_i}^{-1}, \ \beta_{ij}, \gamma_{ij} \ \varepsilon \ A^{p_i}.$$

Let η_{ij}, $\zeta_{ij} \varepsilon A$ be good approximations to β_{ij}, γ_{ij} at p_i and to 1 at $p_{i'}$, $i' = 1, \cdots, t$, $i' \neq i$. Then

$$\xi = \prod_{j=1}^{x_1} (\eta_{ij}\zeta_{ij}\eta_{ij}^{-1}\zeta_{ij}^{-1}) \cdots \prod_{j=1}^{x_t} (\eta_{tj}\zeta_{tj}\eta_{tj}^{-1}\zeta_{tj}^{-1})$$

is a good approximation to α_i at p_i, $i = 1, \cdots, t$.

THEOREM. *Any element of norm 1 in a simple algebra over an algebraic number field is in the commutator group.*

Proof. Let k be the center of the algebra. We prove the theorem in three steps, n being the index of the algebra:

i) $n = 1$, a prime. In this case, the assumption that k is an algebraic number field will not be used. By a remark made before, we may prove the theorem only for a division algebra $D_{(l)}$. Let $\alpha_0 \varepsilon D$ be an element of norm 1. α_0 is contained in a maximal subfield K of D. If D is of characteristic l and K is inseparable over k, α_0 satisfies the equation $\alpha_0{}^l - 1 = 0$ which implies $\alpha_0 = 1$. So, we may assume K is separable over k. Let E be an extension of K, which is normal over k with the Galois group $G(k)$. Let $G(F)$ be an l-Sylow group of $G(k)$ with the fixed field F. Then $(F:k) = m$ is not divisible by l and D_F remains a division algebra. Consider the maximal subfield $K_F = K \times F$ of D_F and let the group of E over K_F be $G(K_F)$. Since $G(K_F)$ is a maximal subgroup in the p-group $G(F)$, it is normal in the same. Hence, K_F is normal and cyclic over F. $\alpha_0 \varepsilon K_F$ is therefore in D'_F by Lemma 2. Applying Lemma 3, we have $\alpha_0{}^m \varepsilon D'$. But Lemma 4 gives $\alpha_0{}^l \varepsilon D'$. Since $(l, m) = 1$, $\alpha_0 \varepsilon D$.

ii) $n = l^s$, $s > 1$. We prove the theorem by induction. Accordingly, assume the theorem has been proved for $n = l^{s-1}$. Let α_0 be an element in $D = D_{(l^s)}$ of norm 1.

Let S_0 be a finite set of finite primes in k, which contains

(a) all finite primes where D is ramified,

(b) all even primes,[5]

(c) a prime which splits completely in the l-component of $k(\zeta_{s+1})$ over k, ζ_{s+1} denoting a primitive l^{s+1}-th root of unity.

For $p \varepsilon S_0$, let K^p be a maximal unramified subfield of D^p, and let θ_p be an element of norm 1, which generates $K^p : K = k_p(\theta_p)$. For example, θ_p may be taken as a primitive $(ql^s - 1)/(q-1)$-th root of unity in K^p, q being the number of residue-classes of p in k. Let $\xi \varepsilon D'$ be a sufficiently good approximation to the elements θ_p for all $p \varepsilon S_0$. Then $\alpha = \xi\alpha_0$ will be a sufficiently good approximation to the elements θ_p, and will generate on account of Lemma 6 maximal subfields in D^p isomorphic to K^p for all $p \varepsilon S_0$. It follows, first of all, that $K = k(\alpha)$ is a maximal subfield of D.

Now, let E, F, $G(k)$, $G(F)$, $G(K_F)$ have the same significance as in i). Passing a composition series of $G(F)$ through $G(K_F)$, we see K_F has a subfield Ω_1, which is cyclic of degree l over F. We are now in the division algebra D_F. Let $\nu_1 = N_{K_F \Omega_1}(\alpha)$.

We shall denote the extension of S_0 in F by S_0 again. Let S^* be the set of all primes in F, which are not relatively prime to ν_1 in Ω_1. Since $N_{\Omega_1F}(\nu_1)=1$ and Ω_1 is unramified of degree l at the primes in S_0, $S^*\cap S=0$. Finally, let $\mathfrak{q}\notin S_0+S^*$ be a prime which remains prime in Ω_1 and in the l-component of $F(\zeta_s)$ over F.

---

[5] S. Wang, "On Grunwald's theorem," to appear in the *Annals of Mathematics*.

By (c), Ω_1 is independent with $F(\zeta_{s+1})$ over F. Applying Corollary 2 in [4] with $\bar{C}=\Omega_1$, we construct a cyclic extension Γ of degree l^{s+1} over F such that

(A) $(\Gamma^{\mathfrak{p}}:F_{\mathfrak{p}})=l^{s+1}$ for $\mathfrak{p}\in S_0$,

(B) $\Gamma^{\mathfrak{p}}=F_{\mathfrak{p}}$ for $\mathfrak{p}\in S^*+\mathfrak{q}$,

(C) if Γ is ramified at any $\mathfrak{p}$ outside $S+S^*+\mathfrak{q}$, $\mathfrak{p}$ remains prime in Ω_1.

Let D_1 be the centralizator of Ω_1 in D_F. It is a division algebra of degree l^{s-1} over Ω_1, and is isomorphic to the division algebra in $D_F\times\Omega_1$. So, if $\mathfrak{p}_\infty$ is a real prime in F, where D_F is ramified, then D_1 is ramified over Ω_1 at $\mathfrak{P}_\infty\mid\mathfrak{p}_\infty$, if and only if $\mathfrak{P}_\infty$ is real, i. e., if and only if Ω_1 is unramified over F at $\mathfrak{p}_\infty$. Let S_∞ be the set of all real primes in F, where D_F is ramified but where Ω_1 is unramified.

Let Ω^* be the subfield of Γ of degree l over F. Let $\bar{S}$ be the set of all primes in F outside $S_0+S^*+\mathfrak{q}$, where Ω^* is ramified over F. Put $S=S_0+S^*+\bar{S}+\mathfrak{q}+S_\infty$.

Let Γ_1 be the subfield of Γ of degree l^s over F. For $\mathfrak{p}\in S$, we define $\Gamma^{*\mathfrak{p}}$ and $(\alpha_{\mathfrak{p}},\Gamma^{*\mathfrak{p}}/F_{\mathfrak{p}})$ as follows:[6]

(a₁) for $\mathfrak{p}\in S_0+S^*+\bar{S}+\mathfrak{q}$, $\Gamma^{*\mathfrak{p}}=\Gamma_1{}^{\mathfrak{p}}$ and

$$(\alpha_{\mathfrak{p}},\Gamma^{*\mathfrak{p}}/F_{\mathfrak{p}})=(\alpha_{\mathfrak{p}},\Gamma_1{}^{\mathfrak{p}}/F_{\mathfrak{p}}),$$

(b₁) for $\mathfrak{p}\in S_\infty$, $\Gamma^{*\mathfrak{p}}$ is complex and $(\alpha_{\mathfrak{p}},\Gamma^{*\mathfrak{p}}/F_{\mathfrak{p}})$ is defined in the only way possible.

Write $S'=S_0+S^*+\bar{S}+\mathfrak{q}$. When $l=2$ and when $F(\zeta_{s+1})$ is not cyclic over F, we have, by Lemma 3 in [4],

$$\prod_{\mathfrak{p}\in S'}((\sec 2\pi/2^{t+1})^{2^{s-1}},\Gamma_1{}^{\mathfrak{p}}/F_{\mathfrak{p}})=1. \tag{5}$$

---

[6] This is the upper part of the orthodox norm residue symbol. The lower part, consisting of a horizontal line and the prime $\mathfrak{p}$, is omitted for printing convenience.

Since $(\sec 2\pi/2^{t+1})^{2^{s-1}}$ is totally positive, (5) implies that

$$\prod_{\mathfrak{p}\,\varepsilon\, S}((\sec 2\pi/2^{t+1})^{2^{s-1}}, \Gamma^{*\mathfrak{p}}/F_{\mathfrak{p}}) = 1$$

and that condition (d) in the theorem in [4] is satisfied. Applying that theorem with $\bar{C} = \Omega_1$, $C = \Omega$, we construct a cyclic extension Γ^* of degree l^s over F such that

(A_1) $\Omega^* \subset \Gamma^*$,

(B_1) the p-adic completion of Γ^* is $\Gamma^{*\mathfrak{p}}$ for $\mathfrak{p}\,\varepsilon\, S$,

(C_1) if Γ^* is ramified at any $\mathfrak{p}$ outside S, $\mathfrak{p}$ remains prime in Ω_1.

Consider the field $K_1 = \Omega_1 \times \Gamma^*$. This is a direct product, because $\mathfrak{q}$ remains prime in Ω_1, but splits completely in Γ^* by (B) and (a_1). We contend that ν_1 is a norm everywhere in K_1. In fact, at all primes extending those in S_∞, $\nu_1 = N_{D_1\Omega_1}(\alpha)$ is a norm in D_1 and is therefore positive. At all primes extending those in S^*, K_1 splits completely over Ω_1. Since ν_1 is a unit at all finite primes not in S^*, it is a norm everywhere in K_1 except possibly at the finite primes $\mathfrak{P}$ where K_1 is ramified over Ω_1. Let $\mathfrak{p}$ be the prime in F divisible by $\mathfrak{P}$. Then $\mathfrak{p}\,\varepsilon\, S_0 + \bar{S} +$ the primes in (C_1). By the construction of K, Γ, and Γ^*, $\mathfrak{p}$ remains prime in Ω_1. We have

$$(\nu_1, K_1{}^{\mathfrak{P}}/\Omega_{1\mathfrak{P}}) = (N_{\Omega_1{}^{\mathfrak{p}}F_{\mathfrak{p}}}(\nu_1), \Gamma^{*\mathfrak{p}}/F_{\mathfrak{p}}) = (1, \Gamma^{*\mathfrak{p}}/F_{\mathfrak{p}}) = 1.$$

Consequently, ν_1 is a norm everywhere in K_1, and there exists an element $\beta_1\,\varepsilon\, K$ such that $N_{K_1\Omega_1}(\beta_1) = \nu_1$.

Consider the algebra $A = M_{(l)} \times D_F$ over F. It contains the algebra $A_1 = M_{(l)} \times D_1$ over Ω_1. Since K_1 clearly splits D_1 and is of degree l^s over Ω_1, it is a maximal subfield of A_1. Consider the element

$$\gamma_1 = \begin{pmatrix} 1 & & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & & \alpha \end{pmatrix}$$

in A_1. We have

$$N_{A_1\Omega_1}(\gamma_1) = N_{D_1\Omega_1}(\alpha) = \nu_1 = N_{K_1\Omega_1}(\beta_1) = N_{A_1\Omega_1}(\beta_1),$$

i. e., $N_{A_1\Omega_1}(\gamma_1) = N_{A_1\Omega_1}(\beta_1)$. By the hypothesis of induction, γ_1 differs from β_1 only by an element in $A'_1 \subset A'$.

Let us assume for a moment that it had already been proved that $\beta_1\,\varepsilon\, A'$.

We would infer then $\gamma_1 \varepsilon A'$. Lemma 1 would give $\alpha \varepsilon D_F$. By the same argument as in i), we would conclude $\alpha \varepsilon D'$. So, $\alpha_0 \varepsilon D'$ and the theorem in the case of prime power indices would have been proved.

To prove that β_1 is actually in A', consider the centralizator A^* of Ω^* in A. It contains of course the element β_1. Now, it is easy to construct a field of degree l^{s-1} over Ω^*, which splits D_F.[7] So, Ω^* is isomorphic to a subfield Ω_2 of D_F of degree l over F. Let D_2 be the centralizator of Ω_2 in D_F. Then that of Ω_2 in A is $A_2 = M_{(l)} \times D_2$. Since A_2 is conjugate to A^* in A, A_2 contains an element β_2 conjugate to β_1. To prove $\beta_1 \varepsilon A'$, it suffices to prove $\beta_2 \varepsilon A'$.

Let $\nu_2 = N_{A_2\Omega_2}(\beta_2)$. Let S^{**} be the set of all primes in F, which are not relatively prime to ν_2 in Ω_2. Let S'_∞ be the set of all real primes in F, where D_F is ramified. Put $S = S_0 + S^{**} + S'_\infty + \bar{S}$.

We are going to construct a cyclic extension K_2 of degree l^{s+1} over F, applying the theorem in [4] with $\tilde{C} = C = \Omega_2$. Since $\mathfrak{q}$ remains prime in the l-component of $F(\zeta_s)$, but splits completely in Ω_2, Ω_2 is independent with $F(\zeta_s)$ over F.

For $\mathfrak{p} \varepsilon S$, we define $K_2^{\mathfrak{p}}$ and $(\alpha_{\mathfrak{p}}, K_2^{\mathfrak{p}}/F_{\mathfrak{p}})$ as follows:

(a$_2$) for $\mathfrak{p} \varepsilon S_0 + \bar{S}$, $K_2^{\mathfrak{p}} = \Gamma^{\mathfrak{p}}$ and $(\alpha_{\mathfrak{p}}, K_2^{\mathfrak{p}}/F_{\mathfrak{p}}) = (\alpha_{\mathfrak{p}}, \Gamma^{\mathfrak{p}}/F_{\mathfrak{p}})$,

(b$_2$) for $\mathfrak{p} \varepsilon S^{**}$, $K_2^{\mathfrak{p}} = F_{\mathfrak{p}}$,

(c$_2$) for $\mathfrak{p} \varepsilon S'_\infty$, $K_2^{\mathfrak{p}}$ is complex and $(\alpha_{\mathfrak{p}}, K_2^{\mathfrak{p}}/F_{\mathfrak{p}})$ is defined in the only way possible.

Since $\Omega_2^{\mathfrak{p}} = F_{\mathfrak{p}}$ for $\mathfrak{p} \varepsilon S^{**}$, $\Omega_2^{\mathfrak{p}} \subset K_2^{\mathfrak{p}}$ for $\mathfrak{p} \varepsilon S^{**}$. As before, we prove that, when $l = 2$ and when $F(\zeta_s)$ is not cyclic over F,

$$\prod_{\mathfrak{p} \varepsilon S} ((\sec 2\pi/2^{t+1})^{2^s}, K_2^{\mathfrak{p}}/F_{\mathfrak{p}}) = 1,$$

and that condition (d) is satisfied. Consequently, there exists a cyclic extension K_2 of degree l^{s+1} over F having the following properties:

(A$_2$) $\Omega_2 \subset K_2$,

(B$_2$) the $\mathfrak{p}$-adic completion of K_2 is $K_2^{\mathfrak{p}}$ for $\mathfrak{p} \varepsilon S$,

(C$_2$) if K_2 is ramified at any prime $\mathfrak{p}$ outside S, $\mathfrak{p}$ remains prime in Ω_2.

Again, we prove that there exists an element γ_2 in K_2 such that $N_{K_2\Omega_2}(\gamma_2) = \nu_2$.

---

[7] For example, let $\eta \varepsilon \Omega^*$ be totally negative and divisible exactly by the first power of $\mathfrak{P}$ for $\mathfrak{P} \varepsilon S_0$. Then $\Omega^*(\eta^{1/a})$, where $a = l^{s-1}$, splits D_F.

Since K_2 splits D_F and therefore D_2, it is a maximal subfield of A_2 and therefore of A. We have $N_{A_2\Omega_2}(\gamma_2) = N_{K_2\Omega_2}(\gamma_2) = \nu_2 = N_{A_2\Omega_2}(\beta_2)$, i. e., $N_{A_2\Omega_2}(\gamma_2) = N_{A_2\Omega_2}(\beta_2)$. By the hypothesis of induction β_2 differs from γ_2 only by an element in $A'_2 \subset A'$. Now,

$$N_{AF}(\gamma_2) = N_{\Omega_2 F}(\nu_2) = N_{AF}(\beta_2) = N_{AF}(\beta_1) = N_{\Omega_1 F}(\nu_1) = N_{D_F F}(\alpha) = 1,$$

i. e., $N_{AF}(\gamma_2) = 1$. But γ_2 is contained in the cyclic maximal subfield K_2 of A. So, γ_2 is a commutator in A and $\beta_2 \varepsilon A'$.

iii) $n = l_1^{s_1} l_2^{s_2} \cdots l_r^{s_r}$, general. Let F_i be an extension of degree m_i of k such that $(m_i, l_i) = 1$, and that D_{F_i} is of index $l_i^{s_i}$ over F_i. Let $\alpha_0 \varepsilon D$ be of norm 1. Then $\alpha_0 \varepsilon D'_{F_i}$. By Lemma 3, $\alpha_0^{m_i} \varepsilon D'$. But, by Lemma 4, $\alpha_0^n \varepsilon D'$. Since $(m_1, m_2, \cdots, m_r, n) = 1$, $\alpha_0 \varepsilon D'$. The theorem is thus completely proved.

If, in iii), the fields F_i are chosen such that $(m_i, l_i^{s_i}) = l_i^{s_i - 1}$ and that D_{F_i} is of index l, we obtain $\alpha_0^{n/(l_1 l_2 \cdots l_r)} \varepsilon D'$. Since i) was proved for general algebras, we have the following

COROLLARY. *Let A be any simple algebra. Then, $N\alpha = 1$ in A implies $\alpha^{n/(l_1 \cdots l_r)} \varepsilon A'$, $n = l_1^{s_1} l_2^{s_2} \cdots l_r^{s_r}$ being the index of A. In particular, if the index of A is square free, any element of norm 1 in A is in A'.*

PRINCETON UNIVERSITY.

ON THE REPRESENTATION OF SURFACES.*

By Lamberto Cesari.

1. Let S be any continuous surface of xyz-space and

$$S: x = x(u,v),\ y = y(u,v),\ z = z(u,v),\ (u,v) \in Q \equiv (0,1,0,1), \tag{1}$$

any representation of S upon the fundamental square Q. Let us call

$$\begin{aligned} \Phi_1:&\quad y = y(u,v), && z = z(u,v), \\ \Phi_2:&\quad z = z(u,v), && x = x(u,v), && (u,v) \in Q, \\ \Phi_3:&\quad x = x(u,v), && y = y(u,v), \end{aligned} \tag{2}$$

the relative plane transformations which we get by projecting S upon the three coordinate planes. T. Radó introduced the notion of a plane transformation essentially of bounded variation (eBV) and this concept is equivalent to the notion of a transformation of bounded variation (BVC) introduced by the author, as T. Radó has proved recently [T. R. 1, 3]. In addition the author proved [L. C. 2] that *the surface S has finite Lebesgue area if and only if the three plane transformations* (2) *are BVC* (*or eBV*).

Each plane transformation BVC (or eBV)

$$\Phi: \quad x = x(u,v), \qquad y = y(u,v), \qquad (u,v) \in Q,$$

admits, almost everywhere (a. e.) in Q, an (absolute) generalized Jacobian $J(u,v)$ according to the equivalent definitions that T. Radó and the author have introduced [T. R. 1, L. C. 1, 3]. Such a Jacobian is L-integrable on Q. If the functions $x(u,v)$, $y(u,v)$ have first partial derivatives or, at least, an asymptotic regular differential, a. e. in a set $M \subset Q$, then we have [T. R. 1, L. C. 3] a. e. in M

$$J(u,v) = |\partial(x,y)/\partial(u,v)|.$$

that is, the generalized Jacobian coincides, a. e. in M, with the ordinary Jacobian.

If the surface S has finite Lebesgue area, then the three transformations (2) are BVC and therefore have generalized Jacobians $J_1(u,v)$, $J_2(u,v)$, $J_3(u,v)$ a. e. in Q and these are L-integrable functions in Q. In addition we have [T. R. 1, L. C. 3]

---

* Received March 18, 1949.

$$(3) \qquad L(S) \geq \int\int_Q [J_1^2(u,v) + J_2^2(u,v) + J_3^2(u,v)]^{1/2}\, du\, dv,$$

that is the Lebesgue area is always greater than or equal to the classical integral calculated with generalized Jacobians.

T. Radó and the author [T. R. 1, L. C. 3] introduced equivalent notions of an absolutely continuous plane transformation (*eAC*, or *ACC*, according to the nomenclature of T. Radó) and showed, at the same time and independently, that *in relation* (3) *the equality sign holds if and only if the three plane transformations* (2) *are eAC* (*or ACC*).

Now the two following problems, that may be termed representation problems [see T. Radó 1, V, 2, 19, p. 475], arise:

I) Given a surface S of finite Lebesgue area, does there exist such a representation (1) for which equality holds in relation (3), that is, for which the Lebesgue area is given by the classical integral calculated with generalized Jacobians?

II) Given a surface S of finite Lebesgue area, does there exist a representation (1) for which the functions $x(u,v)$, $y(u,v)$, $z(u,v)$ have ordinary first partial derivatives a. e. in Q and for which equality holds in relation (3), that is, the Lebesgue area is given by the classical integral calculated with ordinary Jacobians?

In recent papers [L. C. 5, 6, 7] the author gave an affirmative answer to the problem I. In the present paper an affirmative answer is given to the problem II. This problem has been called to the attention of the writer in the course of a correspondence with J. W. T. Youngs.

2. A partial affirmative answer to the problems I and II was first given by E. J. McShane [3] for saddle surfaces and then, more generally, by C. B. Morrey [1] for open nondegenerate surfaces. The theorem of C. B. Morrey, for which the author gave a direct proof by means of the direct method of Calculus of Variations [L. C. 4], asserts the following:

Each continuous, open nondegenerate surface S of finite Lebesgue area admits a representation (1) *such that*:

a) *the functions $x(u,v)$, $y(u,v)$, $z(u,v)$ are of bounded variation and absolutely continuous in the sense of Tonelli in Q*;

b) *the first partial derivatives $x_u, x_v, \cdots, z_v$, that exist a. e. in Q as a consequence of* a), *are L^2-integrable functions in Q*;

c) *a. e. in Q we have $E = G$, $F = 0$, where $E = x_u^2 + y_u^2 + z_u^2$, $G = x_v^2 + y_v^2 + z_v^2$, $F = x_u x_v + y_u y_v + z_u z_v$*;

d) *the Lebesgue area* $L(S)$ *is given by the classical integral:*

$$L(S)=\int\int_Q [(\partial(x,y)/\partial(u,v))^2+\cdot\cdot\cdot]^{1/2}\,du\,dv$$
$$=\int\int_Q (EG-F^2)^{\frac{1}{2}}du\,dv=\tfrac{1}{2}\int\int_Q (E+G)\,du\,dv.$$

3. A point P of Q is said to be an *exceptional point* [L. C. 5] for the representation (1) of the surface S if a neighborhood $U(P)$ of P exists such that the surface defined by (1) upon $U(P)$ has zero Lebesgue area. A set $I\subset Q$ is said to be an *exceptional set* [L. C. 5] if I is open (relative to Q) and all points of I are exceptional. A set $M\subset Q$ is said to be a *proper set* [L. C. 5] if its complementary set $M-Q$ is exceptional.

Let $f(x)$, $a\leq x\leq b$, be a function, continuous in (a,b), and $E\subset(a,b)$ a closed set. We shall define the total variation $V(f,E,\alpha,\beta)$ of $f(x)$ on the set E in the interval (α,β), $a\leq\alpha<\beta\leq b$, as the least upper bound of

$$\sum_{i=1}^{n-1}|f(x_i)-f(x_{i+1})|$$

for each set $\alpha\leq x_1<x_2<\cdot\cdot\cdot<x_n\leq\beta$ of points of E. If in (α,β) there is no point of E or only one, we put $V(f,E,\alpha,\beta)=0$. We shall say that $f(x)$ is of *bounded variation* V upon E in (a,b) if $V(f,E,a,b)<+\infty$ [Cf. S. Saks 1, p. 221]. We shall say that $f(x)$ is *absolutely continuous* V upon the set E in (a,b) if, for any $\epsilon>0$ there is a $\delta>0$ such that, for each set of intervals (α_i,β_i), $i=1,2,\cdot\cdot\cdot,n$, of (a,b) of which no two have common points and $\sum_{i=1}^{n}(\beta_i-\alpha_i)<\delta$, we have

$$\sum_{i=1}^{n}V(f,E,\alpha_i,\beta_i)<\epsilon.$$

Let $f(x,y)$ be a continuous function in the fundamental square

$$Q=[0\leq x\leq 1,0\leq y\leq 1]$$

and $M\subset Q$ a closed point set. For each $\bar{x}$, $0\leq\bar{x}\leq 1$, let us call $M(\bar{x})$ the linear closed set (possibly empty) of the points $(\bar{x},y)$, $0\leq y\leq 1$, of M. The linear set $M(\bar{y})$ has an analogous definition.

For each $\bar{x}$, $0\leq\bar{x}\leq 1$, let $V_y(f,M,Q,\bar{x})$ denote the total variation of the function of y alone, $f(\bar{x},y)$, $0\leq y\leq 1$, on the closed set $M(\bar{x})$ in the interval $0\leq y\leq 1$. $V_x(f,M,Q,\bar{y})$ has an analogous definition. We shall say that $f(x,y)$ is *of bounded variation* V *on the set* M *in the square* Q [L. C. 5], if $V(f,M,Q,x)$, $0\leq x\leq 1$, which is a function of x alone, is finite a. e. and L-integrable, and if the corresponding fact holds for $V_x(f,M,Q,y)$, $0\leq y\leq 1$. We shall say that $f(x,y)$ is *absolutely continuous*

V on the set M in the square Q [L. C. 5], if $f(x,y)$ is of bounded variation V on the set M in the square Q, if for almost all $\bar{x}$, $0 \leq \bar{x} \leq 1$, the function of y alone, $f(\bar{x},y)$, $0 \leq y \leq 1$, is absolutely continuous V on $M(\bar{x})$ in $0 \leq y \leq 1$, and if the same fact holds when x and y are interchanged. If $f(x,y)$ is continuous in Q and of bounded variation V on a closed set M of Q, then $f(x,y)$ admits first partial asymptotic derivatives a. e. in M and also an asymptotic differential [L. C. 7, p. 306]. If we consider that M may be a totally disconnected closed set, it is evident that, in general, the ordinary first partial derivatives and even the regular asymptotic differential, may not exist.

4. Let S be any continuous surface of finite Lebesgue area. The author proved that there always exists a representation of S upon the fundamental square for which the area is given by the classical integral calculated with generalized Jacobians (§ 1, problem I). More precisely, the author proved the following:

THEOREM A. *Each continuous surface S of finite Lebesgue area admits a representation* (1) *such that:*

a) *the functions $x(u,v)$, $y(u,v)$, $z(u,v)$ are of bounded variation V and absolutely continuous V on a closed set $M \subset Q$, where M is a proper set for the representation* (1);

b) *each of the functions $x(u,v)$, $y(u,v)$, $z(u,v)$, which, as a consequence of* a) *has first partial asymptotic derivatives and an asymptotic differential a. e. in M, also has a regular asymptotic differential and the first partial asymptotic derivatives are L^2-integrable on M;*

c) *a. e. in M we have $E = G$, $F = 0$;*

d) *the generalized Jacobians J_1, J_2, J_3 are zero at every point of $Q - M$ (the exceptional set) and coincide with the ordinary ones a. e. in M (the proper set).*

e) *the Lebesgue area $L(S)$ of S is given by the classical integral*

$$L(S) = \int\int_Q (J_1^2 + J_2^2 + J_3^2)^{\frac{1}{2}}\, du\, dv$$
$$= \int\int_M (EG - F^2)^{\frac{1}{2}} du\, dv = \tfrac{1}{2} \int\int_M (E + G)\, du\, dv.$$

This theorem was proved by the author in recently published papers [L. C. 6, 7] and given in the preliminary note [5] already mentioned.

5. First of all we notice that in the Theorem A, the statement b) can be replaced by the following stronger one:

b′) *the functions* $x(u,v)$, $y(u,v)$, $z(u,v)$ *have ordinary first partial derivatives a. e. in* M, *and therefore also regular asymptotic differentials; and the first partial derivatives are* L^2*-integrable in* M.

This stronger statement was proved by the author in the paper [7] already mentioned, but is not given in the preliminary note [5].

In the present paper we shall deduce from Theorem A, with statement b) replaced by b′), the following Theorem B which gives an affirmative answer to problem II (§ 1).

THEOREM B. *Each continuous surface* S *of finite Lebesgue area admits a representation* (1) *upon the fundamental square* Q *such that*

a*) *the functions* $x(u,v)$, $y(u,v)$, $x(u,v)$ *have first partial derivatives a. e. in* Q;

b) *the derivatives* $x_u, x_v, \cdots, z_v$ *are* L^2*-integrable in* Q;

c) *a. e. in* Q *we have* $E = G$, $F = 0$;

d) *the Lebesgue area* $L(S)$ *of* S *is given by the classical integral*;

e) *the functions* x, y, z *are of bounded variation* V *and absolutely continuous* V *in a closed set* $M \subset Q$, *where* M *is a proper set for the representation* (1);

f) *the first partial derivatives* $x_u, x_v, \cdots, z_v$ *are zero a. e. in the exceptional set* $Q - M$.

Remarks. Theorem B holds for each continuous surface and is similar to the theorem of Morrey **2** for open nondegenerate surfaces, but the conclusion a*) of Theorem B is weaker than the conclusion a) of the theorem of Morrey. We observe that in the theorem of Morrey **2** d) is a consequence of a) and b) [see T. Radó, 1, V, 2, 26, p. 480]. In Theorem B d) is not a consequence of a*) and b). E. J. Mickle [1] has proved that there exists a continuous surface of finite Lebesgue area that has no representation (1) with x, y, z functions of bounded variation in the sense of Tonelli in Q. This important result of Mickle proves that in Theorem B conclusion a*) cannot be replaced by the stronger conclusion a).

Proof of Theorem B.

6. For open nondegenerate surfaces Theorem B, with $M \equiv Q$, coincides with the theorem of C. B. Morrey. Let S be a surface of the *type A* [L. C. 5, 6; *base surface* according to C. B. Morrey (1)]. Then Theorem A holds with the following further conclusions [L. C. 6]:

f) *M consists of the boundary Q^* of Q and of a finite or countable set of disjoint circles $\bar{C}_i \subset Q$, $i = 1, 2, 3, \cdots$, and any limit points of these circles lie on the boundary Q^* of Q;*

g) *on the circles C_i equations* (1) *define closed nondegenerate surfaces S_i.*

By a), b′), c) equations (1) give upon each circle C_i a representation of the surface S_i which satisfies the theorem of C. B. Morrey and also the Theorem B. Therefore we have simply to modify the representation (1) only on the exceptional set $Q - M$. Let us divide Q into 4 equal squares each of side-length 2^{-1} and of these let us choose only those that are, with their boundaries, completely interior to $Q - M$; let us call them $q_1, q_2, \cdots, q_{r_1}$. Let us divide each of the remaining squares into 4 equal squares, each of the side-length 2^{-2} and of these let us choose only those that are completely interior to $Q - M$; let us call them $q_{r_1+1}, \cdots, q_{r_2}$, and so on. We get an infinite sequence of squares $q_1, q_2, \cdots, q_j, \cdots$, which fill $Q - M$ and

$$Q - M = \sum_{j=1}^{\infty} q_j, \qquad | Q - M | = \sum_{j=1}^{\infty} | q_j | .$$

We know [see the Appendix] that on each of these squares q_j, there exists a particular plane transformation of q_j into itself

$$\phi_j \colon\ u = u_j(\alpha, \beta),\ v = v_j(\alpha, \beta),\ (\alpha, \beta) \,\varepsilon\, q_j,$$

such that ϕ_j is continuous, monotone, identical on the boundary q^*_j of q_j and constant on a countable set of disjoint squares $q_{jk} \subset q_j$, $k = 1, 2, 3, \cdots$, which are completely interior to q_j and for which we have

$$\sum_{k=1}^{\infty} | q_{jk} | = | q_j | , \qquad\qquad j = 1, 2, \cdots.$$

Let us now put

$$X(\alpha, \beta) = \begin{cases} x[u_j(\alpha, \beta), v_j(\alpha, \beta)] & \text{if } (\alpha, \beta) \,\varepsilon\, q_j, \quad j = 1, 2, \cdots, \\ x(\alpha, \beta) & \text{if } (\alpha, \beta) \,\varepsilon\, M, \end{cases}$$

and analogously let us define $Y(\alpha, \beta)$ and $Z(\alpha, \beta)$. The equations

$$(4) \qquad S\colon\ x = X(\alpha, \beta),\ y = Y(\alpha, \beta),\ z = Z(\alpha, \beta),\ (\alpha, \beta) \,\varepsilon\, Q,$$

give a new representation of the surface S, which coincides with (1) on M.

In a. e. interior point of the circles C_i, $i = 1, 2, \cdots$, and hence a. e. in M, we have $X_u = x_u, \cdots, Z_v = z_v$. Let (α, β) be an interior point of one of the squares q_{jk}, $k = 1, 2, \cdots$, $j = 1, 2, \cdots$. Then the functions u and v are constant on q_{jk} and therefore the functions X, Y, Z are also constant there which implies $X_u = 0, \cdots, Z_v = 0$. This holds a. e. in $Q - M$. Indeed

$$| Q - M | = \sum_{j=1}^{\infty} | q_j | = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} | q_{jk} | .$$

This completes the verification of statements a*) and f). In addition we have $E = G$, $F = 0$ a. e. in M and $E = G = F = 0$ a. e. in $Q - M$. Therefore c) holds. Statements d) and e) follow from Theorem A.

Theorem B for surfaces of type A is now completely proved.

Note. The procedure that we have used here is related to a simpler observation of J. W. T. Youngs [3; see also T. R. 1, V, 2, 21, p. 476].

7. Let S be a closed nondegenerate surface. Then Theorem A holds with the following further conclusions [L. C. 7, p. 342]:

f′) *M consists of a closed circle $\bar{C}$ completely interior to Q, or of the boundary C^* of circle C and a finite or countable set of disjoint circles $\bar{C}_i \subset C$, $i = 1, 2, 3, \cdots$, where any limit points of these circles lie on C^*;*

g′) *on the circle C, or on the circles C_i, equations* (1) *define closed nondegenerate surfaces S_i.*

We have now only to repeat the procedure of **6**.

8. Let S be any continuous surface. Then Theorem A holds but M is now any closed set and furthermore it may be totally disconnected. Let us point out that, while the first partial asymptotic derivatives on M of the functions x, y, z can be calculated by using only values of these functions on M, the ordinary first partial derivatives on M must be calculated by using values of these functions on $Q - M$ as well. In this general case also we modify on $Q - M$ the given representation (1) of Theorem A, but we must be careful not to lose the existence of the first partial derivatives at points of M where they exist (that is, a. e. in M).

Let $\{P, P'\}$, $\{I, I'\}$ denote the distance between two points, or two sets; let $\delta(I)$ be the diameter of the set I. For each $0 < \delta \leq (2)^{\frac{1}{2}}$ set

$$\omega(\delta) = \max \{S(P), S(P')\}$$

for each pair P, P' of points of Q such that $\{P, P'\} \leq \delta$. Of course $\lim_{\delta \to 0} \omega(\delta) = 0$.

As noted in Section **6** there are squares $p_1, p_2, \cdots, p_n, \cdots$ non overlapping and such that $Q - M = p_1 + p_2 \cdots$. For $n = 1, 2, \cdots$, divide p_n into a finite number of congruent squares $p_{n,1}, p_{n,2}, \cdots, p_{n,i_n}$ such that

$$\omega(\operatorname{diam} p_{n,i}) < \{p_n, M\}^2 \leq \{p_{n,i}, M\}^2, \qquad i = 1, 2, \cdots, i_n.$$

Call these squares $q_1, q_2, \cdots, q_n, \cdots$. Then for $n = 1, 2, \cdots$,

α) q_n and its boundary q^*_n lie in $Q - M$;

β) $\omega(\operatorname{diam} q_n) < \{q_n, M\}^2$;

γ) $Q - M = q_1 + q_2 + \cdots + q_n + \cdots$.

Repeating the procedure of **6**, we get a new representation

$$(4) \qquad S\colon x = X(\alpha, \beta),\ y = Y(\alpha, \beta),\ z = Z(\alpha, \beta),\ (\alpha, \beta) \varepsilon Q,$$

of the surface S. Let us prove that the functions X, Y, Z have first partial derivatives a. e. in M, and precisely at the points of M where the functions x, y, z have them.

Let P' be a point of $Q - M$ and consequently a point of a square q_i. Let $\bar{P}$ be any point of q^*_i. We have

$$|X(P') - X(\bar{P})|,\ |x(P') - x(\bar{P})| \leq \omega(\operatorname{diam} q_i), \qquad X(\bar{P}) = x(\bar{P})$$

and the same relations for Y and Z. Hence

$$(5) \qquad |X(P') - x(P')| \leq 2\omega(\operatorname{diam} q_i), \cdots.$$

Now let $P \equiv (\alpha, \beta)$ be a point of M at which the first partial derivative x_u exists; let $P' \equiv (\alpha + h, \beta)$, $h \neq 0$, be any point of Q on the straight line $v = \beta$.

If $P' \equiv (\alpha + h, \beta)$ is a point of M, then

$$[X(\alpha + h, \beta) - X(\alpha, \beta)]/h = [x(\alpha + h, \beta) - x(\alpha, \beta)]/h.$$

If $P' \equiv (\alpha + h, \beta)$ is not contained in M, then P' lies in a square q_i and we have, from (5) and β),

$$[X(\alpha + h, \beta) - X(\alpha, \beta)]/h = [x(\alpha + h, \beta) - x(\alpha, \beta)]/h$$
$$+ [X(P') - x(P')]/h,$$

$$|[X(P') - x(P')]/h| \leq 2[\omega(\operatorname{diam} q_i)/\{q_i, M\}] < 2\{q_i, M\}$$
$$< 2\{P', M\} \leq 2\{P', P\} = 2|h|.$$

In any case

$$\lim_{h \to 0} [X(\alpha + h, \beta) - X(\alpha, \beta)]/h$$
$$= \lim_{h \to 0} [x(\alpha + h, \beta) - x(\alpha, \beta)]/h = x_u(\alpha, \beta),$$

that is, the first partial derivative X_u exists and $X_u(\alpha, \beta) = x_u(\alpha, \beta)$.

The same holds for all other first partial derivatives of the functions

X, Y, Z. This is sufficient to prove that the representation (4) of the surface S satisfies all the conditions of Theorem B.

Consequence of Theorem B.

9. The notion of integral upon a surface

$$J_S = \int\int_Q F(x, y, z, H_1, H_2, H_3)\, du\, dv$$

as a Lebesgue integral has been studied by E. J. McShane [1, 2, 3] for oriented surfaces S that are given by a representation (1) satisfying conditions a) and b) of 2.

More generally T. Radó [2] studied the same notion under the hypotheses that the functions x, y, z have first partial derivatives a. e. in Q and that the Lebesgue area $L(S)$ is finite and given by the classical integral calculated with ordinary Jacobians.

The notion of the integral J_S for any continuous oriented surface S of finite Lebesgue area given by any representation (1) was introduced recently by the author as a Weierstrass integral. The author [8] proved that

1) J_S is independent of the representation of the oriented surface S;

2) if $\| S_n, S \| \rightarrow 0$ in the sense of Fréchet and if $L(S_n) \rightarrow L(S)$, then $J_{S_n} \rightarrow J_S$.

3) J_S is equal to the Lebesgue integral (6) with generalized or ordinary (relative) Jacobians H_1, H_2, H_3 for each representation (1) for which the area is given by the classical integral. From Theorem A we know that there is always at least one such representation.

From Theorem B that we have just proved, we know that each surface S of finite Lebesgue area has at least one representation for which the functions x, y, z have ordinary partial derivatives a. e. and for which the Lebesgue area is given by the classical integral. Therefore, by choosing such a representation, each integral J_S can always be calculated as a Lebesgue integral (6) with ordinary Jacobians, that is, as one of the integrals studied by T. Radó [2].

Appendix.

There exist continuous plane transformations Φ *of a square* q *into itself, monotone on* q*, identical upon the boundary* q^* *of* q*, constant upon a countable set of non overlapping closed squares* $q_i \subset q$, $i = 1, 2, \cdots$, $\sum | q_i | = | q |$.

α) *Proof.* Let us indicate by E_3 the subdivision of a square into 9 equal squares by means of straight lines parallel to the sides. Let us perform E_3 upon q and let q_0 be the closed central square. Let us perform E_3 upon each of the remaining 8 squares and let $q_{i_1 0}$ be the 8 central squares, and so on. By repeating this procedure indefinitely we have a countable set of closed non-overlapping squares $q_0, q_{i_1 0}, q_{i_1 i_2 0}, \cdots, q_{i_1 i_2 \ldots i_{n-1} 0}, q_{i_1 i_2 \ldots i_n 0}, \cdots, i_1, i_2, \cdots, i_n, \cdots = 1, 2, \cdots, 8$, contained in q (Sierpinski, 1). Let K be the collection of all these squares plus all individual points of q not on such squares. K is an upper semicontinuous decomposition of q into continua separating neither q, nor the plane and therefore (R. L. Moore, 1) there exists a continuous transformation f of K into the closed unit circle C. f is a biunique transformation between C^* and q^*. There exists a transformation ϕ of C into q, biunique, which coincides with f^{-1} upon C^*. The continuous monotone transformation $\Phi = \phi f$ of q into itself is identical on q^*, is constant upon the squares $q_{i_1 i_2 \ldots i_n 0}$ and $\sum_n \sum_{i_1 i_2 \ldots i_n} | q_{i_1 i_2 \ldots i_n 0} | = | q |$.

β) *An elementary construction of a particular transformation* Φ.

Let $p = A_1 A_2 A_3 A_4$ be any convex plane quadrilateral. Let d be the diameter of p, O be the center of gravity of the four vertices (of the mass unity), let $p_0 = A'_1 A'_2 A'_3 A'_4$ be the image of p under the similarity transformation with center O and ratio $1/n$ ($n \geqq 3$ integer). Let us divide the sides of p into three equal parts by means of the points B_i, $i = 1, 2, \cdots, 8$, ordered upon p^* starting from A_1 and in the same order as the points A_i. Let us indicate by E_n the subdivision of p into the 9 quadrilaterals p_0, $p_1 = A_1 B_1 A'_1 B_8$, $p_2 = B_1 B_2 A'_2 A'_1$, $p_3 = B_2 A_2 B_3 A'_2$, $p_4 = B_3 B_4 A'_3 A'_2$, $p_5 = B_4 A_3 B_5 A'_3$, $p_6 = B_5 B_6 A'_4 A'_3$, $p_7 = B_6 A_4 B_7 A'_4$, $p_8 = B_7 B_8 A'_1 A'_4$. It is elementary to prove that the 9 quadrilaterals p_i are all convex and have diameters $d_i \leq (3/4) d$, $i = 0, 1, \cdots, 8$. If p is a square, then the subdivision E_3 considered in α) is a particular case of E_n.

Let q, q' be the squares $q \equiv [0 \leq \alpha \leq 1,\ 0 \leq \beta \leq 1]$, $q' \equiv [0 \leq u \leq 1,\ 0 \leq v \leq 1]$. For each integer $n \geq 3$ we define the transformation Φ_n as follows. We perform the subdivision E_3 upon the square q, n times, as in α, letting the central squares always remain undivided, and analogously we apply the subdivision E_n to the square q', n times, letting the central quadrilaterals always remain undivided. We obtain in q certain squares $q_0, q_{i_1 0}, q_{i_1 i_2 0}, \cdots, q_{i_1 i_2 \ldots i_{n-1} 0}, q_{i_1 i_2 \ldots i_n}$ of side length $1/3, 1/3^2, \cdots, 1/3^n, 1/3^n$ and correspondingly in q' certain convex quadrilaterals $q'_0, q'_{i_1 0}, q'_{i_1 i_2}, \cdots, q'_{i_1 i_2 \ldots i_{n-1} 0}, q'_{i_1 i_2 \ldots i_n}$ of diameters $\leq 2^{\frac{1}{2}}/n$, $1/n(3/4) 2^{\frac{1}{2}}$, $1/n(3/4)^2 2^{\frac{1}{2}}, \cdots, 1/n(3/4)^{n-1} 2^{\frac{1}{2}}$, $(3/4)^n 2^{\frac{1}{2}}$ ($i_1, i_2, \cdots, i_n = 1, 2, \cdots, 8$). If δ_n is the maximum of such diameters then $\delta_n \to 0$ when $n \to \infty$. Let us divide each of the squares and quadrilaterals of q and q' into four triangles by means of the two diagonals and let Φ_n be the continuous biunique transformation of q into q' which is linear in each triangle of q and which makes each triangle of q' correspond to the corresponding triangle of q. Let us observe that we have successively divided each side of q and the corresponding side of q' into three equal parts and each of these parts again into three equal parts and so on, and that Φ_n is linear in each part. Therefore Φ_n is identical on the boundary q^* of q. It is easy to prove that the sequence Φ_n, $n = 1, 2, \cdots$, converges uniformly in q toward a transformation Φ having the required properties.

UNIVERSITY OF BOLOGNA AND INSTITUTE FOR ADVANCED STUDY.

BIBLIOGRAPHY.

L. Cesari:

1. "Sulla quadratura delle superficie in forma parametrica," *Bollettino della Unione Matematica Italiana*, Ser. II, Anno IV (1942), pp. 109-117.
2. "Caratterizzazione analitica delle superficie continue di area finita secondo Lebesgue," *Annali della Scuola Normale Superiore di Pisa*, ser. II, vol. 10 (1941), pp. 253-294; vol. II (1942), pp. 1-42.
3. "Sui fondamenti geometrici dell' integrale classico per l'area delle superficie in forma parametrica," *Memorie della Accademia d'Italia*, vol. 13 (1943), pp. 1323-1481.
4. "Criteri di uguale continuità ed applicazioni alla quadratura delle superficie," *Annali della Scuola Normale Superiore di Pisa*, ser. II, vol. 12 (1943), pp. 61-84.
5. "Rappresentazione quasi conforme delle superficie continue," *Accademia Nazionale dei Lincei, Rendiconti*, ser. VIII, vol. 1 (1946), pp. 509-515.
6. "Sulla rappresentazione delle superficie continue di area finita secondo Lebesgue," *Rendiconti, Istituto Lombardo di Scienze e Lettere*, vol. 79 (1945-46), pp. 15-45.
7. "Parametrizzazione delle superficie continue di area finita secondo Lebesgue," *Annali di Matematica pura e applicata* (4), vol. 26 (1947), pp. 301-374.
8. "La nozione di integrale sopra una superficie in forma parametrica," *Annali della Scuola Normale Superiore di Pisa*, ser. II, vol. 13 (1944), pp. 73-117.

K. Menger:

1. "Zur allgemeinen Kurventheorie," *Fundamenta Mathematicae*, vol. 10 (1927), pp. 96-115, in particular, p. 107.

E. J. Mickle:

1. "Some examples in surface area theory," *Rivista di Matematica dell' Università di Parma*, vol. 1 (1950), Fasc. 1.

E. J. McShane:

1. "On the semicontinuity of integrals in the calculus of variations," *Annals of Mathematics,* vol. 33 (1932), pp. 460-484.
2. "Integrals over surfaces in parametric form," *ibid.,* vol. 34 (1933), pp. 815-838.
3. "Parametrizations of saddle surfaces with applications to the problem of Plateau," *Transactions of the American Mathematical Society,* vol. 35 (1933), pp. 715-733.

R. L. Moore:

1. "Concerning upper semi-continuous collections of continua," *ibid.,* vol. 27 (1925), pp. 416-428.

C. B. Morrey:

1. "An analytic characterization of surfaces of finite Lebesgue area," *American Journal of Mathematics,* vol. 57 (1935), pp. 692-702.

T. Radó:

1. *Length and Area,* American Mathematical Society Colloquium Publications, vol. 30, 1948.
2. "On the semicontinuity of double integrals in the parametric form," *Transactions of the American Mathematical Society,* vol. 51 (1942), pp. 336-361.
3. "Two-dimensional concepts of bounded variation and absolute continuity," *Duke Mathematical Journal,* vol. 14 (1947), pp. 587-608.

S. Saks:

1. *Theory of the Integral,* Warsaw, 1937.

W. Sierpinski:

1. "Sur une courbe cantorienne qui contient une image biunique et continue de toute courbe donné," *Comptes Rendus,* vol. 162 (1916), pp. 629-632.

J. W. T. Youngs:

1. "Curves and surfaces," *American Mathematical Monthly,* vol. 51 (1944), pp. 1-11.
2. "The topological theory of Fréchet surfaces," *Annals of Mathematics,* vol. 45 (1944), pp. 753-785.
3. "On surfaces of class K_1," *Bulletin of the American Mathematical Society,* vol. 51 (1945), pp. 669-673.

ON NON-LINEAR DIFFERENTIAL EQUATIONS OF FIRST ORDER.*

By PHILIP HARTMAN and AUREL WINTNER.

It was shown in [7] and [1] that, under suitable conditions on the coefficient function $f = f(x)$ of the differential equation $y'' + f(x)y = 0$, one can conclude the existence of at least one solution $y = y(x)$ representable as a Laplace-Stieltjes transform of a non-decreasing function,

$$(1) \qquad y(x) = \int_0^\infty e^{-tx} d\phi(t), \qquad (d\phi(t) \geqq 0).$$

The first part of the present paper deals with the question of existence of solutions representable in the form (1) in the case of differential equations which, in contrast to the type mentioned before, are of first order but non-linear. In the particular case of a Ricatti differential equation, such a result was obtained in [8]. The general results to be proved below admit of various applications, including one to an implicit equation $f(x, y) = 0$. The unrestricted nature of the solution of the differential equation, that is, the existence of a solution $y = y(x)$ in the large (i. e., on some half-line, say $0 < x < \infty$) is, of course, part of the problem considered.

The second part of the paper deals with conditions which assure the non-existence of unrestricted solutions (and, what is more, with the existence of a constant $L > 0$ having the property that no solution can be defined on an interval whose length exceeds L). If $f(x, y)$ is a continuous function on the (x, y)-plane and $y = y(x)$ is a solution of the differential equation $y' = f(x, y)$ on an interval $a < x < b$, which cannot be continued onto any interval $a < x < b + \epsilon$, where $\epsilon > 0$, then $|y(x)| \to \infty$ as $x \to b - 0$; cf. [4]. The relation between a solution $y = y(x)$ and a corresponding number b will be considered for differential equations $y' = f(x, y)$, which possess no unrestricted solutions.

Part I.

1. A function $y = y(x)$ defined on an interval is called completely monotone there if it possesses derivatives of arbitrarily high order which satisfy

$$(2) \qquad (-1)^n y^{(n)}(x) \geqq 0 \text{ for } n = 0, 1, \cdots$$

---

* Received May 20, 1949.

and for all values of x on this interval. According to the Hausdorff-Bernstein theorem, a function $y(x)$ possesses a representation (1) for $0 < x < \infty$ in terms of a monotone, but not necessarily bounded, function $\phi(t)$ on the half-line $t \geqq 0$ if and only if $y(x)$ is completely monotone on the half-line $0 < x < \infty$. Thus the discussion of existence of solutions (1) of a differential equation is reduced to a discussion of the existence of completely monotone solutions.

(I) *Let* $F(x,y)$, $G(x,y) > 0$ *be defined and of class* C^∞ *for* $x > 0$, $y > 0$ *and let their partial derivatives satisfy*

$$(-1)^j \partial^{j+k} F/\partial x^j \partial y^k \geqq 0 \text{ for } j,k = 0,1,2,\cdots \tag{3}$$

and

$$(-1)^{j+1} \partial^{j+k} G/\partial x^j\, \partial y^k \geqq 0 \text{ for } (j,k) \neq (0,0). \tag{4}$$

Then every solution $y = y(x)$ (> 0) *of*

$$F + Gy' = 0 \tag{5}$$

is completely monotone on any interval on which it exists.

It may be remarked that (5) can be written in the form $F/G + y' = 0$ and (I) can then be applied to the case where F, G are replaced by F/G, 1 respectively. The wording of (I) and this application of (I) lead to different criteria for (5) to possess only completely monotone solutions (that is, it is possible for the functions F, G to satisfy conditions (3), (4) whether or not the functions $F/G, 1$ satisfy them). This will be shown in **3**.

(I) leads to the following result concerning implicit equations:

Corollary 1_I. *Let* $f(x,y)$ *be defined and of class* C^∞ *for* $x > 0$ *and* $y > 0$ *and let its partial derivatives satisfy*

$$(-1)^{j+1} \partial^{j+k} f/\partial x^j \partial y^k \geqq 0 \text{ for } (j,k) \neq (0,0),\ (0,1), \text{ while } \partial f/\partial y > 0. \tag{6}$$

In addition, let

$$-\infty \leqq \lim_{y\to+0} f(x,y) < 0 \text{ and } 0 < \lim_{y\to\infty} f(x,y) \leqq \infty \tag{7}$$

for every fixed $x > 0$. *Then there exists a unique solution* $y = y(x)$ *of the implicit equation*

$$f(x,y) = 0 \tag{8}$$

for $0 < x < \infty$, *and this* $y = y(x)$ *is completely monotone for* $0 < x < \infty$.

If the conditions of (I) are modified slightly, then certain solutions of (5) are representable in the form (1), where $\phi(t)$ is monotone and bounded on $0 \leqq t < \infty$.

COROLLARY 2_I. *Let* $F(x,y)$, $G(x,y) > 0$ *be defined and of class* C^∞ *for* $x \geqq 0$, $y \geqq 0$ *and such as to satisfy* (3), (4), *respectively. In addition, let*

$$F(x,0) \equiv 0 \text{ for } 0 \leqq x < \infty. \tag{9}$$

Then the solution of (5) *determined by an initial condition*

$$y(0) = y_0 \ (\geqq 0) \tag{10}$$

exists on the entire half-line $0 \leqq x < \infty$ *and is completely monotone there (in particular, if* $y_0 > 0$, *then* $y(x) > 0$ *for all* $x \geqq 0$).

It may be remarked that if $y = y(x)$ is any solution of (5), under the conditions of Corollary 2_I, then there exists a number $a \geqq 0$ such that the definition of $y = y(x)$ can be extended over a half-line $a < x < \infty$ and $y(x)$ is completely monotone there. Consequently, $y(x)$ has a representation of the form (1) valid for $a < x < \infty$, where $\phi(t)$ is monotone (not necessarily bounded) on the half-line $0 < t < \infty$.

Another type of modification of (I) leads to cases in which a solution of (5) is not completely monotone but is the primitive function of a completely monotone function (that is, $y'(x)$ is completely monotone).

(II) *Let* $F(x,y)$, $G(x,y) > 0$ *be defined and of class* C^∞ *for* $x > 0$, $y > 0$ *and let their partial derivatives satisfy*

$$(-1)^{j+k}\partial^{j+k}F/\partial x^j \partial y^k \geqq 0 \text{ for } j, k = 0, 1, 2, \cdots, \tag{11}$$

and

$$(-1)^{j+k+1}\partial^{j+k}G/\partial x^j \partial y^k \geqq 0 \text{ for } (j,k) \neq (0,0). \tag{12}$$

Then every solution $y = y(x)$ (> 0) *of*

$$F - Gy' = 0 \tag{13}$$

possesses a completely monotone derivative $y'(x)$ *on the interval on which* $y(x)$ *is defined.*

In contrast to the situation in (I), the assertion of (II) is in its most inclusive form when $G \equiv 1$; that is, if F, G satisfy the conditions of (I) and F, G are replaced by $F/G, 1$ respectively, then (II) is still applicable (but the converse is obviously false).

Corresponding to the Corollary 1_I, there results the following corollary of (II):

COROLLARY 1_{II}. *If* $f(y)$, *where* $0 < y < \infty$, *is an increasing function of class* C^∞ *satisfying either*

$$(14) \qquad (-1)^k d^k f/dy^k \geqq 0 \text{ for } k = 2, 3, \cdots$$

or, more generally,

$$(15) \qquad (-1)^k d^k (df/dy)^{-1}/dy^k \geqq 0 \text{ for } k = 0, 1, 2, \cdots$$

(*which means that* $(df/dy)^{-1}$ *is completely monotone*), *then the inverse function* $y = y(x)$ *of* $x = f(y)$ *possesses a completely monotone first derivative* $y'(x)$.

Also, if the conditions of (II) are modified so as to assure the existence of unrestricted solutions of (13), an analogue of Corollary 2_I is obtained.

COROLLARY 2_{II}. *Let* $F(x, y)$, $G(x, y) > 0$ *be defined and of class* C^∞ *for* $x \geqq 0$, $y > 0$ *and let their partial derivatives satisfy* (11), (12), *respectively. In addition, let there exist positive continuous functions* $\lambda = \lambda(s)$ *and* $\phi = \phi(s)$ *defined for* $0 < s < \infty$ *satisfying*

$$(16) \qquad F(x, y)/G(x, y) \leqq \lambda(x)\phi(y)$$

and

$$(17) \qquad \int^\infty \lambda(s)ds < \infty \text{ and } \int^\infty ds/\phi(s) = \infty.$$

Then the solution $y = y(x)$ *of* (13) *determined by the initial condition*

$$(18) \qquad y(0) = y_0 \ (> 0)$$

exists on the entire half line $0 \leqq x < \infty$ *and possesses a completely monotone first derivative* $y'(x)$.

Remarks similar to those following the Corollary 2_I are applicable to the last corollary also.

2. *Proof of* (I). Let $y = y(x)$ be a solution of (5) on an x-interval S; so that $y(x) > 0$ on S (since F and G are defined only for $x > 0$, $y > 0$). Let (2_n) denote the inequality in (2) for a fixed n. Thus $y(x) > 0$ implies (2_0). Also, (5) and $F \geqq 0$, $G \geqq 0$ imply (2_1). Suppose that (2_0), (2_1), $\cdots$, (2_n) hold on S. It must be verified that (2_{n+1}) holds on S.

The chain rule for differentiation shows that, if $n > 0$, the n-th derivative of $F(x, y(x))$ consists of a sum of terms of the type

$$(\partial^{j+k}F/\partial x^j\partial y^k)(dy/dx)^\alpha(d^2y/dx^2)^\beta\cdots(d^ny/dx^n)^\nu \tag{19}$$

where

$$j+\alpha+2\beta+\cdots+n\nu=n. \tag{20}$$

This is clear for the case $n=1$, since $dF(x,y(x))/dx=\partial F/\partial x+(\partial F/\partial y)\times(dy/dx)$. Assume that the statement is correct for a given $n\geqq 1$. Then a differentiation of (19) shows that the statement is correct for $n+1$. For the derivative of the first factor in (19) is

$$\partial^{j+k+1}F/\partial x^{j+1}\partial y^k+(\partial^{j+k}F/\partial x^j\partial y^{k+1})(dy/dx);$$

so that the sums (20) belonging to the corresponding two terms are $(j+1)+\alpha+2\beta+\cdots=n+1$ and $j+(\alpha+1)+2\beta+\cdots=n+1$, respectively. The derivative of a factor, say $(d^ey/dx^e)^\epsilon$, is $\epsilon(d^ey/dx^e)^{\epsilon-1}(d^{e+1}y/dx^{e+1})$, so that the sum (20) corresponding to this term is $j+\alpha+2\beta+\cdots+[e(\epsilon-1)+(e+1)]+\cdots$ or $j+\alpha+2\beta+\cdots+(e\epsilon+1)+\cdots=n+1$. Hence, by (3) and the induction hypothesis on $(2_0),\cdots,(2_n)$, the inequality $(-1)^nd^nF(x,y(x))/dx^n\geqq 0$ holds on S.

Similarly, the chain and product rules show that the n-th derivative of $G(x,y(x))y'(x)$ is a sum of $Gy^{(n+1)}$ and terms of the type

$$(\partial^{j+k}G/\partial x^j\partial y^k)(dy/dx)^\alpha(d^2y/dx^2)^\beta\cdots(d^ny/dx^n)^\nu$$

where $j+\alpha+2\beta+\cdots+n\nu=n+1$. Hence, by (3) and the induction hypothesis,

$$(-1)^n(d^n(G(x,y(x))y'(x))/dx^n-G(x,y(x))d^{n+1}y/dx^{n+1})\geqq 0.$$

If (5) is differentiated n times, the result can be written in the form $Gd^{n+1}y/dx^{n+1}=-Q$, where $(-1)^nQ\geqq 0$. Since $G>0$, it follows that (2_{n+1}) holds. This completes the proof of (I).

Proof of Corollary 1_I. The condition $\partial f/\partial y>0$ and (7) imply the existence of a unique solution $y=y(x)>0$ of (8) for $0<x<\infty$. Clearly, $y=y(x)$ is of class C^∞ and satisfies the differential equation (5), where $F(x,y)=\partial f/\partial x$ and $G(x,y)=\partial f/\partial y$. Since $G=\partial y/\partial y>0$, it only remains to show that conditions (3), (4) are satisfied. Since

$$(-1)^j\partial^{j+k}F/\partial x^j\partial y^k=(-1)^j\partial^{j+1+k}f/\partial x^{j+1}\partial y^k\geqq 0$$

and

$$(-1)^{j+1}\partial^{j+k}G/\partial x^j\partial y^k=(-1)^{j\mp 1}\partial^{j+1+k}f/\partial x^j\partial y^{k+1}\geqq 0$$

by (6), Corollary 1_I follows from (I).

Proof of Corollary 2_I. In virtue of (I), the Corollary 2_I will be proved if it is shown that a solution of (5) defined on some interval $0 \leqq a \leqq x \leqq b$ can be extended over the half line $a \leqq x < \infty$. The smoothness (C^∞) of F and G implies that if such an extension is possible, then it is necessarily unique by virtue of "local uniqueness." That is, if $x_0 \geqq 0$, $y_0 \geqq 0$, then there exist some number $\delta > 0$ and a unique solution of (5) on $x_0 \leqq x \leqq x_0 + \delta$ satisfying $y(x_0) = y_0$. This is also true for $y_0 = 0$; in fact, by (9), such a solution is given by $y(x) \equiv 0$, while the smoothness of F and G assures the uniqueness of this solution. In particular, if a solution $y = y(x)$ of (5) vanishes for some value of $x = x_0 \geqq 0$, then $y \equiv 0$.

Consider an arbitrary solution $y = y(x)$ ($\geqq 0$) of (5) on some interval $0 \leqq a \leqq x \leqq b$. If $y(x)$ vanishes at some point of this interval, it vanishes identically and possesses the continuation $y(x) \equiv 0$ for $a \leqq x < \infty$. If $y(x) > 0$ on the interval, then $F \geqq 0$ and $G > 0$ imply that $y'(x) \leqq 0$, by (5). Thus, $y(x)$ possesses either an extension over the half-line $a \leqq x < \infty$ or over some interval $a \leqq x < x_0$, where $(x, y(x))$ tends to a boundary point of the domain of definition of F and G as $x \rightarrow x_0$ (that is, $y(x) \rightarrow 0$ holds as $x \rightarrow x_0$). In the latter case, if $y(x_0)$ is defined to be zero, the function $y = y(x)$ becomes a solution of (5) on the closed interval $a \leqq x \leqq x_0$, and $y(x_0) = 0$. But then $y(x) \equiv 0$, which contradicts $y(x) > 0$ on $a \leqq x \leqq b$. Consequently, $y(x)$ possesses a continuation over the half-line $a \leqq x < \infty$ (and $y(x)$ does not vanish on this half-line). This completes the proof of Corollary 2_I.

Proof of (II). The proof of this statement is similar to that of Theorem (I) and can be omitted.

Proof of Corollary 1_{II}. If (14) holds, then $df/dy > 0$ for $0 < y < \infty$, by virtue of the strict monotony of f. For if $df/dy = 0$ when $y = y_0 > 0$, then $df/dy = 0$ when $0 < y \leqq y_0$, since $df/dy \geqq 0$ and $d^2f/d^2y \geqq 0$, by (14). But this contradicts the assumption that $f(y)$ is strictly monotone. The assumption (15) is meaningless unless $df/dy > 0$ for $0 < y < \infty$. Consequently in either case, the inverse function $y = y(x)$ of $x = f(y)$ possesses a continuous first derivative and satisfies the differential equation $1 - (df/dy)y' = 0$. Corollary 1_{II} now follows from (II) by considering F, G in (13) to be $(df/dy)^{-1}$, 1, respectively. In the general case, where (15) is assumed, (11) reduces to (15) if $j = 0$, and to $0 \geqq 0$ if $j > 0$. Thus, (II) is applicable and the proof of Corollary 1_{II} is complete.

Proof of Corollary 2_{II}. In virtue of (II), it is sufficient to show that if $y = y(x)$ is a solution of (13) defined on some interval $a \leqq x \leqq b$, then the definition of $y(x)$ can be extended over the half-line $a \leqq x < \infty$. But this follows from a known result [5], p. 451. [Strictly speaking, [5] becomes applicable only if $F(x, y)/G(x, y)$ is defined for $0 \leqq x < \infty$ and $-\infty < y < \infty$. But if it is noticed that (13) implies that $y'(x) \geqq 0$, so that $y(x) \geqq y(a) > 0$ for all $x > a$ for which $y(x)$ is defined, it is clear that the domain of definition of F/G, namely, $x \geqq 0$, $y > 0$, is sufficient for the purposes at hand.]

3. It was mentioned above that if (5) is written in the form

$$F/G + y' = 0 \tag{5'}$$

and (I) is applied to (5′), where F, G are replaced by F/G, 1, respectively, then the resulting criteria for (5) and/or (5′) to have only completely monotone solutions neither imply nor are implied by (I) itself. That is to say, (3) and (4) neither imply nor are implied by

$$(-1)^j \partial^{j+k}(F/G)/\partial x^j \partial y^k \geqq 0 \quad \text{for } j, k = 0, 1, \cdots. \tag{3'}$$

In order to see this, first let $F(x, y) = e^y$ and $G(x, y) = y$; so that $G > 0$ and (3), (4) are satisfied when $y > 0$ but (3′) is not (e. g., if $j = 0$ and $k = 1$, then $\partial(F/G)/\partial y = d(y^{-1}e^y)/dy = e^y y^{-1}(1 - y^{-1}) < 0$ if $y < 1$). Next, let $F(x, y) \equiv 1$ and $G(x, y) = e^{-y}$. Then $G > 0$, and (3′) is satisfied (since $F/G = e^y$) but (4) is not (e. g., if $j = 0$ and $k = 2$, then $(-1)\partial^2 G/\partial y^2 = -e^{-y} < 0$).

It may be observed that, although (14) implies (15), the converse is not true. For example, $f(y) = e^y$ satisfies (15) but not (14).

Part II.

4. This part will be concerned with the existence or non-existence of unrestricted solutions of differential equations, that is, of solutions existing on some half-line $a \leqq x < \infty$. The assumption (9) in Corollary 2_I and the assumption involving (16), (17) in Corollary 2_{II} merely play the rôle of assuring that all solutions of (5) and (13), respectively, are unrestricted. These conditions can be replaced by other types of conditions which assure the existence of only some unrestricted solutions. In this direction, the following will be proved:

(III) *Let $f(x)$ be a continuous function defined for large positive x such that*

$$f(x) \geqq 0 \tag{21}$$

and that

$$r'' + f(x)r = 0 \tag{22}$$

is non-oscillatory (e. g., let

$$\limsup x^2 f(x) < 1/4 \tag{23}$$

as $x \to \infty$). *Let* $F(x, y)$ *be defined and continuous for* $x > 0$, $y > 0$ *and satisfy*

$$0 \leqq F(x, y) < y^2 + f(x) \tag{24}$$

on some half-strip $X \leqq x < \infty$, $0 < y \leqq \epsilon$, *where* $X > 0$, $\epsilon > 0$. *Then the differential equation*

$$F(x, y) + y' = 0 \tag{25}$$

possesses some unrestricted solutions; in particular, if X *is sufficiently large and if* $y = y(x)$ *is a solution of* (25) *and satisfies the initial condition*

$$y(x_0) = y_0, \text{ where } x_0 \geqq X \text{ and } y_0 \geqq \epsilon, \tag{26}$$

then all continuations of $y = y(x)$ *can be extended over the half-line* $x_0 \leqq x < \infty$.

This theorem can be combined with (I) to assure the existence of solutions of (5) which are representable in the form (1) on some half-line $x_0 \leqq x < \infty$.

The differential equation (22) is said to be non-oscillatory if one (hence every) solution $r = r(x) \not\equiv 0$ of (22) has only a finite number of zeros on the half-line $X \leqq x < \infty$ (for sufficiently large X). That (23) is sufficient for (22) to be non-oscillatory is a result of A. Kneser [2], p. 415. The last part of (III) involving "continuations" of a solution is needed, since the conditions on $F(x, y)$ do not imply the local uniqueness of solutions of (25).

The theorem (III) has an analogue involving the assumption that (22) is oscillatory and the conclusion that no solution of (25) is unrestricted.

(IV) *Let* $f(x)$, *a continuous function defined for* $0 \leqq x < \infty$, *be such as to make* (22) *oscillatory (e. g., let*

$$\liminf x^2 f(x) > 1/4 \tag{27}$$

as $x \to \infty$). *Let* $F(x, y)$ *be defined and continuous for* $x > 0$, $-\infty < y < \infty$ *and let it satisfy*

$$F(x, y) > y^2 + f(x) \qquad (X \leqq x < \infty, -\infty < y < \infty). \tag{28}$$

Then no solution of (25) *is unrestricted.*

The fact that (27) implies that (22) is oscillatory (i. e., not non-oscillatory) was proved by Kneser [2], p. 415. One can modify (IV) so as to assure the existence of a constant $L > 0$ with the property that no solution of (25) is defined over an interval $(0 <)\ x_0 < x < x_0 + L$ of length L.

COROLLARY 1_{IV}. *Suppose that* $f(x)$, $F(x, y)$ *satisfy the conditions of* (IV) *and that the distance between successive zeros of any non-trivial solution* $r = r(x)$ *of* (22) *does not exceed a fixed* $L > 0$ (*e. g., let*

$$f(x) \geqq \pi^2/L^2 \tag{29}$$

for $0 \leqq x < \infty$). *If* $y = y(x)$ *is a solution of* (25) *satisfying an initial condition* $y(x_0) = y_0$, *then to every continuation of* $y(x)$ *there corresponds a number* b *with the property that*

$$y(x) \to -\infty \text{ as } x \to b - 0 \tag{30}$$

and $x_0 < b < x_0 + L$.

It may be remarked that if (25) is replaced by

$$y' = F(x, y) \tag{31}$$

then (IV) remains valid; the same is true of its Corollary 1_{IV} except that (30) must be replaced by

$$y(x) \to \infty \text{ as } x \to b - 0. \tag{32}$$

Corollary 1_{IV} can be generalized as follows:

COROLLARY 2_{IV}. *Let* $f(x)$, $g(y)$ *be defined and continuous for* $0 \leqq x < \infty$, $-\infty < y < \infty$, *respectively, and suppose that*

$$f(x) \geqq c > 0, \qquad g(y) \geqq 0 \tag{33}$$

and

$$L = \int_{-\infty}^{\infty} dy/(g(y) + c) < \infty. \tag{34}$$

Let $F(x, y)$ *be defined and continuous for* $0 \leqq x < \infty$, $-\infty < y < \infty$ *and such as to satisfy*

$$F(x, y) \geqq f(x) + g(y). \tag{35}$$

Then the conclusions of Corollary 1_{IV} *are valid.*

Conditions for the non-existence of unrestricted solutions of (25) can take a very different form. An application of the theorem in [6], p. 554 for the differential equation $dx/dy = 1/F(x, y)$ implies the following result:

(V) *Let* $F(x,y)$ *be defined and continuous for* $0 \leqq x < \infty$, $0 \leqq y < \infty$ *and let there exist two positive, continuous functions* $\lambda(s)$, $\phi(s)$ *on the half-line* $0 < s < \infty$ *satisfying*

$$F(x,y) \geqq \lambda(x)\phi(y) \tag{36}$$

and

$$\int^{\infty} ds/\lambda(s) < \infty \text{ and } \int^{\infty} \phi(s)\,ds = \infty. \tag{37}$$

Then no solution of (31) *is unrestricted; furthermore, there belongs to every number* $b > 0$ *at least one solution* $y = y(x) = y(x;b)$ *of* (31) *on* $0 \leqq x < b$ *satisfying* (32).

The solutions $y = y(x;b)$ of this theorem can be made unique by adapting the conditions of the theorems of [3] to the situation at hand; cf. [3], p. 131.

5. *Proof of* (III). Let $r = r(x) \not\equiv 0$ be a solution (22). Then, since (22) is supposed to be non-oscillatory, $r(x)$ does not vanish for sufficiently large x, say for $x > X$. It can be supposed that $r(x) > 0$ for large x (for otherwise r can be replaced by $-r$). Then (21) implies that the graph of $r = r(x)$ is convex downwards for large x. Hence $r > 0$, $r' \geqq 0$, $r'' \leqq 0$, since r has no zeros for large x. Let $z = r'/r$. Then z satisfies the Ricatti differential equation

$$z' + z^2 + f(x) = 0 \qquad (z = r'/r) \tag{38}$$

for $x > X$. Clearly, $z > 0$ and $z' \leqq 0$. Hence, $z(x)$ tends to a limit $z(\infty) \geqq 0$ as $x \to \infty$. In fact, $z(\infty) = 0$; for if $z(\infty) > 0$, then (21) and (38) show that $z' < -$const., where const. > 0, for large x, which contradicts $z \geqq 0$ for $x > X$.

For the given $\epsilon > 0$, let X be so large that $0 < z(x) < \epsilon$ if $x \geqq X$. Consider a solution $y = y(x)$ of (25) and the initial condition (26), so that $y(x_0) \geqq \epsilon > z(x_0)$.

The first inequality in (24) and the differential equation (25) show that $y'(x) \leqq 0$. Hence every continuation $y(x)$ for increasing x can be extended over the half-line $x_0 \leqq x < \infty$, unless there exists a number $b > x_0$ such that $y(x)$ is defined for $x_0 \leqq x < b$ and $y(x) \to 0$ as $x \to b - 0$. It will be shown that such a number b cannot exist; in fact, $y(x) > z(x) > 0$, where $x \geqq x_0$, holds for every continuation of $y(x)$. For suppose, if possible, that such is not the case for some continuation. Then

$$\epsilon \geqq y(x) > z(x) \text{ for } x_1 \leqq x < x_2 \text{ but } y(x_2) = z(x_2)$$

holds for a pair of numbers x_1 ($\geqq x_0$) and x_2. According to (25) and (38),

$$(39) \qquad y'(x) - z'(x) = z^2(x) + f(x) - F(x, y(x)).$$

Hence,

$$y'(x_2) - z'(x_2) = y^2(x_2) + f(x_2) - F(x_2, y(x_2) > 0,$$

by (24). Consequently $0 = y(x_2) - z(x_2) > y(x_2 - \delta) - z(x_2 - \delta) > 0$ for sufficiently small $\delta > 0$. This contradiction shows that every continuation of $y(x)$ is unrestricted, which completes the proof of (III).

Proof of (IV). The assumption that (22) is oscillatory implies that if $z = z(x)$ is a solution of (38) on some interval ($X \leqq$) $x_0 \leqq x \leqq x_1$, then, for a suitable choice of x_1, it follows that $z(x) \rightarrow -\infty$ as $x \rightarrow x_1 - 0$. In order to see this, let $r(x)$ be the solution of (22) determined by the initial conditions $r(x_0) = 1$ and $r'(x_0) = z(x_0)$. Then the local uniqueness of solutions of (38) implies that $z(x) = r'(x)/r(x)$ holds on any interval on which $z(x)$ exists. If x_1 is chosen to be the first zero of $r(x)$ to the right of x_0, then $z(x) \rightarrow -\infty$ as $x \rightarrow x_1 - 0$.

Suppose, if possible, that (25) has an unrestricted solution $y = y(x)$ on ($X \leqq$) $x_0 < x < \infty$. Consider the solution $z = z(x)$ of (38) determined by the initial condition $z(x_0) = y(x_0)$. It will be shown that

$$y(x) < z(x) \text{ for } x_0 < x < x_1,$$

which will imply that $y(x) \rightarrow -\infty$ as $x \rightarrow b - 0$ for some $b \leqq x_1$. This will contradict the assumption that $y(x)$ is unrestricted.

Since (25) and (38) imply (39), it follows from the assumption (28) that $y' - z' < 0$ at $x = x_0$; and so $y(x) < z(x)$ if $x > x_0$ is sufficiently near x_0. Suppose that the relation in the last formula line fails to hold. Let x_2 ($> x_0$) be the first value of x (to the right of x_0) at which it does not hold, so that $y(x_2) = z(x_2)$. Then it is seen from (39) and (28) that $y' - z' < 0$ at $x = x_2$. Since $y(x_2) - z(x_2) = 0$, it follows that $y(x) - z(x) > 0$ if x ($< x_2$) is sufficiently near x_2. This is obviously a contradiction and completes the proof of (IV).

Proof of Corollary 1_{IV}. The proof of this assertion is a consequence of the proof just completed, since the assumption on the zeros of a solution $r = r(x)$ of (22) implies that the numbers x_0, x_1 (of the last proof) satisfy $x_1 - x_0 < L$.

The parenthetical part of Corollary 1_{IV}, that which asserts that (29) is sufficient to assure that the distance between successive zeros of a solution

$r = r(x)$ of (22) does not exceed L, follows from Sturm's comparison theorem if the harmonic oscillator, $r'' + (\pi^2/L^2)r = 0$, is used as a "minorant" for (22).

Proof of Corollary 2_{IV}. Let $y = y(x)$ be a solution of (25) on some interval $(0 \leqq)\ x_0 \leqq x < b$. Then (33), (35) and the differential equation (25) imply that $y'(x) < 0$ and that $1 = -y'/F(x, y(x)) \leqq -y'/(f(x) + g(y(x))$, hence $1 \leqq -y'/(g(y(x)) + c)$. Let the latter inequality be integrated from $x = x_0$ to $x = b$. If the monotone function $y = y(x)$ is introduced as a new variable in the integral on the right, one obtains

$$b - x_0 \leqq \int_{\alpha}^{\beta} dy/(g(y) + c), \text{ where } \alpha = y(b - 0), \beta = y(x_0).$$

It follows therefore from (34) that $b - x_0 < L$.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] P. Hartman and A. Wintner, "On the Laplace-Fourier transcendents," *American Journal of Mathematics*, vol. 71 (1949), pp. 367-372.

[2] A. Kneser, "Untersuchungen über die reellen Nullstellen der Integrale linearer Differentialgleichungen," *Mathematische Annalen*, vol. 42 (1893), pp. 409-435.

[3] A. Wintner, "Asymptotic equilibria," *American Journal of Mathematics*, vol. 68 (1946), pp. 125-132.

[4] ———, "The infinities in the non-local existence problem of ordinary differential equations," *ibid.*, vol. 68 (1946), pp. 173-178.

[5] ———, "An Abelian lemma concerning asymptotic equilibria," *ibid.*, vol. 68 (1946), pp. 451-454.

[6] ———, "Asymptotic integration constants," *ibid.*, vol. 68 (1946), pp. 553-559.

[7] ———, "On the Laplace-Fourier transcendents occurring in mathematical physics," *ibid.*, vol. 69 (1947), pp. 87-98.

[8] ———, "A priori Laplace transformations of linear differential equations," *ibid.*, vol. 71 (1949), pp. 587-594.

ON THE SPECTRA OF TOEPLITZ'S MATRICES.*

By Philip Hartman and Aurel Wintner.

---

1. If f_n, where $n = 0, \pm 1, \cdots$, is a sequence of complex numbers satisfying

$$f_n = \bar{f}_{-n} \text{ and } \sum_{n=-\infty}^{\infty} | f_n |^2 < \infty, \tag{1}$$

let L, T, H denote the Laurent, Toeplitz, Hankel matrices defined by $L = (f_{n-m})$, where $n, m = 0, \pm 1, \cdots$, and $T = (f_{n-m})$ and $H = (f_{n+m+1})$, where $n, m = 0, 1, \cdots$, respectively. Thus an L-matrix consists of two (identical) T-matrices on its main diagonal and two (complex-conjugate) H-matrices on its counter-diagonal.

Let x^+ denote a vector $(x_0, x_1, \cdots)$ and x^- a vector $(\cdots, x_{-2}, x_{-1})$, finally (x^-, x^+) a vector $(\cdots, x_{-1}, x_0, x_1, \cdots)$. When no misunderstanding is possible, x will be used to represent a vector of any of the types x^+, x^-, (x^-, x^+). The symbol $| x |$ will denote the length of x. The second condition in (1) implies that $y^+ = Tx^+$, $y^+ = Hx^+$ and $(y^-, y^+) = L(x^-, x^+)$ are defined (without being necessarily of finite length) whenever the x-vector is of finite length.

The problem of determining the spectrum of L was solved by Toeplitz ([3], [4]; cf. [2], pp. 152-155) and is comparatively simple. According to the Fischer-Riesz theorem, there exists a (unique, real-valued) function $f(\theta)$ of class (L^2) on $(0, 2\pi)$, having $\sum f_n e^{in\theta}$ as its Fourier series,

$$f(\theta) \sim \sum_{n=-\infty}^{\infty} f_n e^{in\theta}. \tag{2}$$

In terms of this $f(\theta)$, the situation can be described as follows: The matrix L is bounded in Hilbert's sense if and only if the function $f(\theta)$ is essentially bounded, that is, $| f(\theta) | \leqq$ const. for almost all θ on $(0, 2\pi)$. Furthermore, the number λ belongs to the spectrum, S_L, of L if and only if the measure of the set of θ-values on $(0, 2\pi)$ satisfying $| f(\theta) - \lambda | < \epsilon$ is of positive measure for every fixed $\epsilon > 0$. Similarly, the point spectrum, P_L, contains a given λ if and only if the measure of the set of θ-values satisfying $f(\theta) - \lambda = 0$ is of positive measure. In other words, S_L is identical with the spectrum of the distribution function of $f(\theta)$, and P_L with the set of

---

* Received June 3, 1949.

discontinuity points of this distribution function. In fact, the resolvent of L is the Laurent matrix belonging to the function $(\lambda - f(\theta))^{-1}$. This is readily verified, since Parseval's relation implies that $y = Lx$ is equivalent to $f(\theta)x(\theta) = y(\theta)$, where $x(\theta) \sim \Sigma\, x_n e^{in\theta}$ and $y(\theta) \sim \Sigma\, y_n e^{in\theta}$.

All of this depends on the formal circumstance that the product of two L-matrices is again an L-matrix (and that this multiplication is simply isomorphic to the multiplication of the respective functions f). Correspondingly, because no such rule of multiplication holds for T-matrices, we could find nothing in the literature on the location of their spectra.

The purpose of the following considerations is to fill somewhat this gap. The results are anything but of a final nature. They reach far enough to show that the spectral situation for T-matrices is quite different from that for L-matrices. The case of the H-matrices, which is again different, will be considered at the end of the paper.

2. According to Toeplitz,

(I) *T is bounded if and only if L is; or, equivalently, if and only if the function* (2) *is bounded* (*almost everywhere*).

This is easily verified as follows: Let $L_j = (f_{n-m})$, where $n, m = j, j+1, \cdots$, be a section (in the lower right-hand corner) of L. Then L is bounded if and only if L_j is bounded for every fixed j and the sequence of numbers representing the norms of the matrices L_j is bounded for $j = 0, \pm 1, \cdots$. Since L_j is identical with T (except that its domain is a "different" Hilbert space), the assertion of (I) follows.

If S_L, S_T denote the spectra of T, L, respectively, (I) might suggest that S_T is identical with S_L. In fact, the proof of (I) implies that the least and greatest points of S_T coincide with those of S_L. Further circumstantial evidence results by observing that, on the one hand, the finite sections ${}_jT = (f_{n-m})$, where $n, m = 0, 1, \cdots, j$, and ${}_jL = (f_{n-m})$, where $n, m = -j, -j+1, \cdots, j$, of T and L, respectively, satisfy ${}_{2j+1}T = {}_jL$ and that, on the other hand, a necessary condition for λ to be in S_T [in S_L] is the existence of a sequence of numbers $\lambda_1, \lambda_2, \cdots$, having the property that λ_j is an eigenvalue of ${}_jT$ [of ${}_jL$], and $\lambda_j \rightarrow \lambda$ as $j \rightarrow \infty$. The difficulty is that this necessary condition is not sufficient, as is shown by an example of Toeplitz representing the matrix of the quadratic form $\Sigma\, 2x_{2n}x_{2n+1}$. Actually this example is neither a T- nor an L-matrix; however, the results below (cf. (II)) imply that there exist an L-matrix and a sequence of numbers $\lambda_1, \lambda_2, \cdots$ satisfying the above conditions, while the limit point λ is not in S_L.

It turns out that the above guess is wrong. In this direction, it will be proved that

(II) *S_T contains S_L but need not be identical with S_L.*

In view of the preceding remarks, the first part of (II) implies the following criterion:

(III) *If the function* (2) *is continuous on* $(0, 2\pi)$, *then S_T is identical with S_L* (*which is an interval unless* $f = \text{const.}$)

3. A little more probing into the question seems to indicate that S_T is *always* an interval (not only for a continuous f) unless $f = \text{const.}$; more specifically, that S_T is connected with *the spectrum of the distribution function of that* (*bounded*) *harmonic function* $u(r, \theta)$ on the unit circle $r < 1$ which, in terms of (1), satisfies $u(r, \theta) \to f(\theta)$ as $r \to 1 - 0$ for almost all θ. Correspondingly, it seems that T, in contrast to L, cannot possess a point spectrum unless $f = \text{const.}$ (almost everywhere).

We were unable to prove either of these conjectures. All that will be proved in their favor is contained in the following facts:

(IV) *If T is not a constant multiple of the unit matrix, and if λ is in the point spectrum of L, then λ is not in the point spectrum of T. Moreover, the least and greatest points of S_T* (*if they are distinct, i. e., if* $f \neq \text{const.}$) *are not in the point spectrum of T.*

It can be expected that the above conjectures are provable in the case of "smooth" functions (1). In this direction, it will be shown that

(*) *T cannot have a point spectrum if*

$$F(z) = \sum_{n=-\infty}^{+\infty} f_n z^n \not\equiv \text{const.} \qquad (f_n = \bar{f}_{-n}) \tag{3}$$

is a rational function of z. (Needless to say, since (2) is of class (L^2), the rational function (3) has no pole on the unit circle $|z| = 1$.)

4. The proofs proceed as follows.

Proof of (II). Let λ be a point of S_L. It is no loss of generality to suppose that $\lambda = 0$. For, if $I = (\delta_{nm})$, where $n, m = 0, \pm 1, \pm, \cdots$ or $n, m = 0, 1, \cdots$, is the unit matrix, then $L - \lambda I$ is also an L-matrix, and $T - \lambda I$ is the corresponding T-matrix. Since $\lambda = 0$ is in S_L, there belongs to every $\epsilon > 0$ a point $x = x_\epsilon = (\cdots, x_{-1}, x_0, x_1, \cdots)$ of Hilbert's space satisfying $|x| = 1$ and $|Lx| < \epsilon$. But $|x^j| \to |x|$ and $|L(0, x^j)| \to |Lx|$ as

$j \to \infty$, if $x^j = (x_{-j}, x_{-j+1}, \cdots)$. Hence, $|x^j| > 1 - \epsilon$ and $|Tx^j| < 2\epsilon$ for sufficiently large j. Consequently, $\lambda = 0$ is in the spectrum of T. This proves the first part of (II).

The second part of (II) follows from the first part of (IV). For example, if (1) is a step-function which assumes only a finite set of distinct values $\lambda_1, \lambda_2, \cdots, \lambda_n$, then S_L is a pure point spectrum consisting of the eigenvalues $\lambda = \lambda_j$, where $j = 1, \cdots, n$ (and each eigenvalue has an infinite multiplicity). By (II), the point λ_j is in S_T. But $\lambda = \lambda_j$ is not in the point spectrum of T, by (IV). Hence, $\lambda = \lambda_j$ is not an isolated point of S_T. Since $\lambda = \lambda_j$ is an isolated point of S_L (if multiplicities are ignored), the second part of (II) follows.

Proof of (IV). If $x^{\pm}$, $y^{\pm}$ are defined as at the beginning of Section 1, it is clear that $(y^-, y^+) = L(0, x^+) = (Hx^+, Tx^+)$, and that Parseval's relation implies the identity $f(\theta)x^+(\theta) = y^+(\theta) + y^-(\theta)$ (for almost all θ). Thus, if λ is in the point spectrum of T and x^+ is a corresponding eigenvector, then

$$y^+(\theta) = \lambda x^+(\theta) \text{ and } (f(\theta) - \lambda)x^+(\theta) = y^-(\theta), \text{ where } y^-(\theta) \sim \sum_{n=1}^{\infty} y_{-n} e^{-in\theta}.$$

Let λ be in the point spectrum of L; it can be supposed that $\lambda = 0$. Then $f(\theta)$ vanishes on a set of positive measure. Suppose, if possible, that λ is in the point spectrum of T and that $x^+ \neq 0$ is an eigenvector satisfying

$$f(\theta)x^+(\theta) = y^-(\theta). \tag{4}$$

Accordingly, $y^-(\theta)$ vanishes on a set of positive measure on $(0, 2\pi)$. It follows therefore from a theorem of F. and M. Riesz, conjectured by Fatou, that $y^-(\theta) \equiv 0$, since $y^-(-\theta)$ is of class (L) and is the boundary function of the power series $\sum y_{-n} z^n$ on the unit circle $|z| < 1$. Thus $f(\theta)x^+(\theta) \equiv 0$, and so $x^+(\theta) = 0$ when $f(\theta) \neq 0$. Since $x^+ \neq 0$, it follows that $x(\theta)$ cannot vanish on a set of positive measure on $(0, 2\pi)$. Hence $f(\theta) \equiv 0$. But then T is the zero matrix, which contradicts the assumption that T is not a constant multiple of the unit matrix. Consequently, $\lambda = 0$ is not in the point spectrum of T. This proves the first part of (IV).

In order to prove the second part of (IV), it can be supposed that $\lambda = 0$ is the least point of S_T, and therefore of S_L. Thus $f(\theta) \geqq 0$ for almost all θ. Suppose that $\lambda = 0$ is in the point spectrum of T and let $x^+ \neq 0$ denote an eigenvector satisfying (4). The Fourier expansions of $x^+(\theta)$, $y^-(\theta)$ show that they are orthogonal,

$$\int_0^{2\pi} \overline{x^+(\theta)} y^-(\theta)\, d\theta = 0; \text{ consequently, } \int_0^{2\pi} f(\theta) |x^+(\theta)|^2 \, d\theta = 0,$$

by (4). Since $f(\theta) \geqq 0$ and $x^+(\theta) \equiv 0$, this leads to the same contradiction as before.

Proof of (*). Since $f_n = \bar{f}_{-n}$, the function (3) satisfies the functional equation $F(1/z) = F(z)$. Let N be the order of the pole, if any, of $F(z)$ at $z = 0$. It will first be shown that $F(z)$ has at least $2N$ zeros on the z-plane.

Only the case $N > 0$ need be considered. The rational function $F(z)$ can be expressed in the form $F(z) = P(z)/z^N Q(z)$, where $P(z)$, $Q(z)$ are (relatively prime) polynomials of degree j, k, respectively, satisfying $P(0) \neq 0$, $Q(0) \neq 0$. It is easy to see that $j \geqq 2N$. In fact, since

$$F(1/z) = z^N P(1/z)/Q(1/z) = z^{N-j+k}\{z^j P(1/z)\}/\{z^k Q(1/z)\},$$

and since $z P(1/z)$, $z^k Q(1/z)$ are polynomials which do not vanish at $z = 0$, it follows from $F(z) = F(1/z)$ that $N - j + k = -N$, hence $j = 2N + k \geqq 2N$.

Suppose, if possible, that $\lambda = 0$ is in the point spectrum of T, and let $x^+ \neq 0$ be an eigenvector satisfying (4). Then, formally,

$$(5) \qquad F(z)X(z) = \sum_{n=1}^{\infty} y_{-n} z^{-n}, \text{ where } X(z) = \sum_{n=0}^{\infty} x_n z^n.$$

Since $X(z)$ is regular for $|z| < 1$, the function $F(z)X(z)$ is meromorphic for $|z| < 1$. On the other hand, (5) shows that $F(z)X(z)$ has a regular analytic continuation for $0 < |z| < \infty$.

If $F(z)$ has no pole at $z = 0$, then (5) implies that $y_{-1} = y_{-2} = \cdots = 0$. Since $F(z) \not\equiv 0$, it follows that $X(z) \equiv 0$. This contradicts $x^+ \neq 0$.

Consequently, it can be supposed that $F(z)$ has at $z = 0$ a pole of order $N \geqq 1$. Then

$$z^N F(z)X(z) = y_{-1} z^{N-1} + y_{-2} z^{N-2} + \cdots + y_{-N}.$$

But since $z^N F(z)$ has at least $2N$ zeros, at least N zeros of $F(z)$ are in the circle $|z| \leqq 1$.

Since the polynomial on the right-hand side in the last formula line has at most $N - 1$ zeros, $X(z)$ has at least one pole in the circle $|z| \leqq 1$. Since $X(z)$ is regular for $|z| < 1$, the pole must be on $|z| = 1$. But this is impossible, since $x^+(\theta) = X(e^{i\theta})$ is of class (L^2) on $(0, 2\pi)$. This contradiction completes the proof of (*).

5. It is easy to show that (II) and (IV) contain the following corollary:

(α) *T cannot be completely continuous unless it is the zero matrix.*

In order to prove this assertion, (α), suppose that (α) is false. Then, for some T distinct from the zero matrix, the set of the cluster points of S_T consists of the single point $\lambda = 0$. It follows therefore from (II.) that S_L consists of a sequence $\lambda_1, \lambda_2, \cdots$ satisfying $\lambda_j \rightarrow 0$ as $j \rightarrow \infty$. Furthermore, not every λ_j is 0, since otherwise L, and therefore the corresponding T, is the zero matrix. Accordingly, some $\lambda_j \neq 0$ is in S_T but is not in the point spectrum of T, by (IV). Hence, T cannot be completely continuous.

It may be mentioned that the assertion of (α) holds for L-matrices also:

(β) *L cannot be completely continuous unless it is the zero matrix.*

In fact, (β) is a corollary of the following assertion:

(β^*) *L cannot have a point spectrum containing an eigenvalue of finite multiplicity.*

In order to prove the latter assertion, (β^*), suppose that λ is a value contained in the point spectrum of L. This will be the case if and only if meas $E_\lambda > 0$, where E_λ denotes that subset of the interval $0 \leqq \theta < 2\pi$ on which $f(\theta) = \lambda$. Correspondingly, it is readily seen from Parseval's relation that $x = (\cdots, x_{-1}, x_0, x_1, \cdots)$ is an eigenvector of L, belonging to the eigenvalue λ, whenever the corresponding function $x(\theta)$ vanishes on the complement of E_λ. Since the linear space of such functions $x(\theta)$ cannot be finite-dimensional, (β^*) follows.

For the Hankel matrices H, defined after (1), the situation is quite different, since (α) and (β) become contrasted by the following fact:

(γ) *H must be completely continuous whenever the w-image of the circle $|z| < 1$ under the mapping*

$$w(z) = \sum_{n=1}^{\infty} f_n z^n \tag{6}$$

is of finite area, i. e., $\sum_{n=1}^{\infty} n\,|f_n|^2 < \infty$.

In fact, a matrix (a_{nm}) must be completely continuous if $\sum\sum |a_{nm}|^2$ is convergent (Hilbert). Since the convergence of this double series is equivalent to the convergence of the simple series $\sum n\,|f_n|^2$ if $a_{nm} = f_{n+m-1}$, the assertion of (γ) follows from the definition of H.

6. In order to make the Hankel matrix $H = (f_{n+m-1})$ *Hermitian,* it will from now on be assumed that $\{f_n\}$ is a *real* sequence; cf. (1). Thus

$$g(\theta) \sim \sum_{n=1}^{\infty} f_n \cos n\theta \quad \text{and} \quad h(\theta) \sim \sum_{n=1}^{\infty} f_n \sin n\theta \tag{7}$$

are the real and imaginary parts of (6) on $|z| = 1$, where $z = e^{i\theta}$.

The following criterion goes back to Toeplitz ([3]; cf. [2], p. 154 and [1], p. 223):

(a) *H must be bounded whenever either of the functions* (7) *is bounded* (*almost everywhere*) *for* $0 \leqq \theta \leqq \pi$.

This criterion, (a), suggests that (γ) in Section **5** admits of the following variant:

(b) *H must be completely continuous whenever either of the functions* (7) *is continuous for* $0 \leqq \theta \leqq \pi$.

Clearly, the two sufficient criteria supplied by (b) are independent of each other, and of the criterion supplied by (γ).

In order to prove (b), suppose, for instance, that the first of the functions (7) is continuous for $0 \leqq \theta \leqq \pi$. Then there exist cosine polynomials $g_1(\theta), g_2(\theta), \cdots$ which tend to $g(\theta)$ uniformly for $0 \leqq \theta \leqq \pi$. On the other hand, the norm of the Hankel matrix H belonging to $g(\theta)$ does not exceed const. $\max | g(\theta) |$, where the const. is an absolute constant (cf. [1], p. 223), and the correspondence between H and $g(\theta)$ is distributive. Consequently, if H_k denotes the Hankel matrix belonging to the trigonometric polynomial $g_k(\theta)$, then, since the uniformity of $g_k(\theta) \rightarrow g(\theta)$ means that $\max | g(\theta) - g_k(\theta) | \rightarrow 0$ as $k \rightarrow \infty$, it follows that the norm of $H - H_k$ tends to 0 as $k \rightarrow \infty$. Hence, H must be completely continuous if every H_k is. Finally, the complete continuity of every H_k follows from (γ) in Section **5**, since the power series (6) belong to H_k, being a rational polynomial, is regular for $| z | \leqq 1$.

7. We do not know what is necessary and sufficient for the boundedness of an Hermitian Hankel matrix. One part of the criterion, supplied by (I) in Section **2** for T-matrices, is true for H-matrices, the other is not. In fact, the situation is as follows:

(i) *H is bounded if, but not only if, the corresponding T* (*or L*) *is.*

First, since f_n is now supposed to be real, (1), (2) and (7) show that $f(\theta) = f_0 + 2g(\theta)$, where f_0 is a constant. It follows therefore from (a) in Section **6** that if $f(\theta)$ is a bounded function, then H is a bounded matrix. Hence the positive part of the last italicized statement, (i), follows from (I) in Section **2**. On the other hand, the negative part of (i) follows from (a) by choosing $f_n = n^{-1}$ in (7), since the function $h(\theta) = \Sigma\, n^{-1} \sin n\theta$ is, but the conjugate function $g(\theta) = \Sigma\, n^{-1} \cos n\theta$ is not, bounded for $0 < \theta < \pi$.

Because of the contrast between (I) and (i), the determination of S_H is even more difficult than that of S_T (cf. (II), (III) and (*) in Sections **2** and **3**). In this direction, and in view of the guesses formulated before (IV), the proof of the following fact is not without interest:

(ii) $\lambda = 0$ *is in the cluster spectrum of every (bounded, Hermitian) Hankel matrix.*

It is understood that by the cluster spectrum of a bounded, Hermitian matrix is meant the λ-set consisting of the continuous spectrum and of the cluster values of the points of the point spectrum, with the proviso that an eigenvalue of infinite multiplicity is considered as a cluster value of the point spectrum. According to Weyl ([5], p. 378), the point $\lambda = 0$ is in the cluster spectrum of a bounded, Hermitian matrix H if and only if there exists in Hilbert's space a sequence of unit vectors $x^0, x^1, x^2, \cdots$ which tend weakly to the zero vector and are transformed by H into vectors $Hx^0, Hx^1, Hx^2, \cdots$ which tend strongly to the zero vector.

It follows that $\lambda = 0$ is sure to be in the cluster spectrum of H if $| He^k | \rightarrow 0$ as $k \rightarrow \infty$, where $e^0, e^1, \cdots$ denotes the sequence of vectors which form the successive columns of the unit matrix. Hence, in order to prove (ii), it is sufficient to ascertain that $| He^k | \rightarrow 0$ is true when H is a Hankel matrix, $H = (f_{n+m-1})$. But it is clear from the definition of e^k that, for such an H,

$$| He^k |^2 = \sum_{n=k}^{\infty} f_n{}^2,$$

and so $| He^k | \rightarrow 0$ follows from (1).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

1. G. H. Hardy, J. E. Littlewood and G. Pólya, *Inequalities*, Cambridge, 1934.
2. F. Riesz, *Les systèmes d'équations linéaires à une infinité d'inconnues*, Paris, 1913.
3. O. Toeplitz, "Zur Theorie der quadratischen Formen von unendlichvielen Veränderlichen," *Göttinger Nachrichten*, 1910, pp. 489-506.
4. ———, "Zur Theorie der quadratischen und bilinearen Formen von unendlichvielen Veränderlichen," *Mathematische Annalen*, vol. 70 (1911), pp. 351-376.
5. H. Weyl, "Ueber beschränkte quadratische Formen, deren Differenz vollstetig ist," *Rendiconti del Circolo Matematico di Palermo*, vol. 27 (1909), pp. 373-392.

ON THE SOLUTIONS OF THE EQUATION OF HEAT CONDUCTION.*

By Philip Hartman and Aurel Wintner.

There are obvious analogies between solutions of Fourier's equation

$$u_{xx} = u_t \tag{1}$$

on the open unit square

$$S: \quad 0 < x < 1,\ 0 < t < 1 \tag{2}$$

and solutions of Laplace's equation

$$u_{xx} + u_{tt} = 0 \tag{3}$$

on the open unit circle

$$C: \quad x^2 + t^2 < 1. \tag{4}$$

These analogies deal with questions of existence, uniqueness, and limits at the boundary points of (2), (4), respectively. The object of this paper is to develop these analogies in a certain direction.

By a solution of (1) on (2) (or (3) on (4)) is meant a function for which u_{xx}, u_t (or u_{xx}, u_{tt}) exist and satisfy (1) (or (3)). In particular, u_x, u_t exist; however, this does not imply the continuity of u. What is true in this regard is that if u is a continuous solution of (3) on some open domain, then u is regular analytic on that domain. The corresponding statement for solutions of (1) is as follows:

(I) *If $u(x,t)$ is a continuous solution of* (1) *on an open domain, then $u(x,t)$ is of class C^∞ on that domain.*

This follows from (II) and (IV) below.

Let K denote the set of points on the three line segments which form the lower and lateral boundaries of S, so that

$$K = I' + J + I'',$$

where I', J, I'' denote the respective segments

$$I': (x = 0, 0 \leqq t < 1), \qquad I'': (x = 1, 0 \leqq t < 1). \tag{5}$$
$$J: (0 < x < 1, t = 0),$$

---

* Received April 9, 1949.

Gevrey ([2], pp. 372-374) has proved the following theorem of uniqueness:

(II) *Let* $u(x, t)$ *be defined on* $S + K$ *and have the following properties: u is uniformly continuous on S, vanishes identically on K, and is a solution of* (1) *on S. Then u vanishes identically.*

Actually, Gevrey's wording of his theorem assumes the continuity of u_x and u_t, but a glance at his proof shows that this pair of additional assumptions is not used at all. Gevrey's proof of (II) depends on a refined, but still very elementary, form of the classical maximum principle, namely, on the following fact:

(III) *Let* $u = u(x, t)$ *be defined on the closure of S and have the properties that it is uniformly continuous and satisfies* (1) *on S. Then u assumes its maximum* (*for the closure of S*) *on K.*

The analogues of radial boundary limits for solutions of (3) on (4) are the 1-dimensional boundary limits,

$$(6_1)\qquad \lim_{x\to+0} u(x, s) = f_1(s), \qquad\qquad (6_2)\qquad \lim_{x\to 1-0} u(x, s) = f_2(s),$$

and

$$(6_3)\qquad \lim_{t\to+0} u(s, t) = f_3(s),$$

for the case of solutions of (1) on (2). These limits (when they exist) form the "boundary function" of $u(x, t)$ on the set K.

Let $\theta(x, t)$ denote the Jacobi ϑ_3-function, with a suitable change of units on the axes of the independent variables,

$$(7)\qquad \theta(x, t) = 1 + 2 \sum_{k=1}^{\infty} \exp(-k^2\pi^2 t) \cos k\pi x.$$

Put

$$(8_1)\qquad u_1(x, t) = -\int_0^t \theta_x(x, t - s) f_1(s)\, ds,$$

$$(8_2)\qquad u_2(x, t) = \int_0^t \theta_x(x - 1, t - s) f_2(s)\, ds,$$

and

$$(8_3)\qquad u_3(x, t) = \tfrac{1}{2} \int_0^1 \{\theta(x - s, t) - \theta(x + s, t)\} f_3(s)\, ds,$$

where $\theta_x = \partial\theta/\partial x$. The existence theorem mentioned above is as follows:

(IV) *Let* $f_1(s)$, $f_2(s)$, $f_3(s)$ *form a uniformly continuous boundary function on* K; *i. e., let each of these functions be continuous for* $0 \leqq s \leqq 1$ *and satisfy* $f_1(0) = f_3(0)$ *and* $f_2(0) = f_3(1)$. *Then* (8_1), (8_2), (8_3) *exist for every* (x, t) *on* S *(as Lebesgue integrals) and their sum,*

$$u(x,t) = u_1(x,t) + u_2(x,t) + u_3(x,t), \tag{9}$$

represents a solution of (1) *on* S, *is uniformly continuous on* S *and satisfies* (6_1)-(6_3) *for* $0 \leqq s \leqq 1$.

This theorem seems to be between the lines of the classical literature but is not explicitly stated and proved. That (9) is a continuous solution of (1) on S, and satisfies (6_1)-(6_3) for $0 < s < 1$, is contained in standard facts; cf., e. g., [7], pp. 111-118. But this does not imply the important point in (IV), namely, the uniform continuity of (9); in this regard, cf. the remarks below concerning the function (10). Actually, the existence of a solution of (1) on S which is uniformly continuous on S and satisfies (6_1)-(6_3) for $0 \leqq s < 1$ follows from the existence theorems in [5], pp. 214-229; [2], pp. 317-342. Furthermore, the representation (9) of this (unique) uniformly continuous solution can be verified by showing that the Green functions, whose existence is proved in [5], [2], reduce to the kernels in (8_1)-(8_3), when the domain in question is S.

A more direct verification of (IV) can be sketched as follows: Simple calculations show the truth of (IV) in the cases in which

$$f_1(s) = 1, \qquad f_2(s) = 1, \qquad f_3(s) = 1$$

for $0 \leqq s \leqq 1$;

$$f_1(s) = 0, \qquad f_2(s) = 1, \qquad f_3(s) = s;$$

for an arbitrary positive integer k,

$$f_1(s) = 0, \qquad f_2(s) = 0, \qquad f_3(s) = \sin k\pi s;$$

finally,

$$f_1(s) = s^k, \qquad f_2(s) = 0, \qquad f_3(s) = 0.$$

These sets (f_1, f_2, f_3), along with those which result if f_1 and f_2 (and correspondingly, s and $1 - s$ in f_1 and f_2) are interchanged, form a Weierstrassian basis of all continuous functions on $K = I' + I'' + J$. The existence of a solution u of (1) on S, uniformly continuous on S and reducing to a given continuous boundary function on K, follows from the maximum principle, (III). The existence of this solution u and its representation (9) can be deduced from an adaptation of the maximum principle for functions having representations (9).

In fact, (9) can be written as

$$u(P) = \int_K G(P, Q) f(Q) dK,$$

where P and Q denote points of S and K, respectively, and $f(Q)$ denotes (f_1, f_2, f_3), a continuous function on K. But $G(P, Q) \geqq 0$ (and equality holds only when Q is a corner of K), by (8_1)-(8_3); cf. (19), (26) and (27) below. Also,

$$\int_K G(P, Q) dK = 1$$

for every P in S; cf. the case $(f_1, f_2, f_3) \equiv (1, 1, 1)$ above. Clearly, the desired maximum principle follows; for $u(P)$ is a weighted average, with non-negative weights and total weight 1, of the boundary function $f(Q)$.

Unfortunately, the assumptions of the uniqueness theorem (II) do not correspond to the physical situation. Correspondingly, since Hadamard's critique of the various types of "well-set" problems on partial differential equations, the following observations seem to be generally known (cf., e. g. [1], pp. 365-366, where the objection is only implicit).

For the *parabolic* differential equation in the plane, the requirement of two-dimensional continuity, in case of a "well-set" boundary value problem, is quite artificial, since there is only *one* sheaf of characteristic curves.

The physical content of this objection is readily realized. In fact, since u, t, x in (1) denote temperature, time and position on a rod, respectively, the data of the boundary value problem, a problem to which the above statements, (II) and (IV), supply theorems of existence and uniqueness, are as follows: The initial temperature $f_3(x)$, at every point of the rod $0 \leqq x \leqq 1$, and the temperatures, $f_1(t)$ and $f_2(t)$, for every later date $(0 < t \leqq 1)$, at both ends, $x = 0$ and $x = 1$, of the rod. But this interpretation makes unnatural indeed the above assumption on the two-dimensional uniform continuity of $u(x, t)$ on S (that is, the two-dimensional continuity on the closure of S). All that is natural to require is the continuity of the temperature distribution (a) *on the rod,* at every *fixed* date t, and (b) *in time,* at every *fixed* point x of the rod. The two "stokers," engaged at $x = 0$ and $x = 1$, seem to need quite an advance knowledge of analytic functions in the real domain if, while working independently of each other and of the past, they succeed in producing two-dimensional continuity by virtue of the one-dimensional assumptions (a), (b).

On the other hand, it is well-known that the omission of the assumption of uniform continuity in (II) invalidates the uniqueness statement. This is shown by the classical "source function," defined by

$$(10)\qquad \phi(x,t)=xt^{-3/2}\exp(-x^2/4t) \text{ if } t>0.$$

It is readily verified that (10) is a solution of (1) on S, that the radial limits (6_1)-(6_3) exist and form a continuous function of the position on K, with $f_1(s)\equiv 0$, $f_2(s)=s^{-3/2}\exp(-1/4s)$ and $f_3(s)\equiv 0$. However, (10) is not the solution supplied by (9), since (10) is not uniformly continuous; in fact, it is not bounded on S. This is seen by considering the function (10) on a parabolic arc $t=x^2$, as $x\to+0$.

Part of these objections can be met by the following existence and uniqueness theorems which, in contrast to (II) and (IV), do not involve the assumption of two-dimensional continuity on the boundary of S:

(V) *Let* $f_1(s)$, $f_2(s)$, $f_3(s)$ *be bounded, measurable functions on* $0<s<1$. *Then* (8_1)-(8_3) *exist for* (x,t) *on* S *(as Lebesgue integrals) and their sum* (9) *is a bounded, continuous solution of* (1) *on the open square* S *and satisfies* (6_1)-(6_3) *almost everywhere.*

(VI) *Conversely, if* $u(x,t)$ *is a bounded, continuous solution of* (1) *on the open square* S, *then the limits* (6_1)-(6_3) *exist almost everywhere and, for* (x,t) *on* S, *the function* $u(x,t)$ *is given by* (9) *in terms of the functions* (8_1)-(8_3).

The condition of two-dimensional uniform continuity in (II) can now be relaxed to that of boundedness. In fact, it follows from (VI), (IV) and (II) that if $u(x,t)$ is a bounded, continuous solution of (1) on S and if the limits (6_1)-(6_3) exist for *every* s on $0<s<1$ and form a uniformly continuous function of the position on K, then $u(x,t)$ is uniformly continuous on the open square S.

The corresponding theorem is true for harmonic functions, bounded on C. However, in the case of harmonic functions, the assumption of "bounded on C" can be relaxed to "half-bounded on C," for example,

$$(11)\qquad u(x,t)\geqq 0.$$

The analogous statement for solutions of (1) on S is false. This is shown by the classical "source" function (10).

This, the impossibility of relaxing the assumption of "boundedness" to "half-boundedness" in (VI), is quite unfortunate, since "half-boundedness" does not require more than an appeal to the *third* law of thermodynamics, whereas the justification of "boundedness" depends on the outcome of an argument about the possible fuel supply of the stokers and their observance of the *first* law.

In order to understand this discrepancy between the behavior of certain solutions of (1) and (3), it is necessary to consider more general solutions of (1) than those given by (V). A well-known theorem (cf., e. g. [3]) states that if $u(x,t)$ is a harmonic function on C and satisfies (11), then $u(x,t)$ is representable as the Stieltjes integral of the Poisson kernel with respect to a bounded non-decreasing function. The analogue of such a Poisson-Stieltjes integral for solutions of (1) would be (9), where

$$(12_1)\qquad u_1(x,t) = -\int_0^t \theta_x(x,t-s)\,dF_1(s),$$

$$(12_2)\qquad u_2(x,t) = \int_0^t \theta_x(x-1,t-s)\,dF_2(s)$$

and

$$(12_3)\qquad u_3(x,t) = \tfrac{1}{2}\int_0^1 \{\theta(x-s,t)-\theta(x+s,t)\}\,dF_3(s),$$

and $F_1(s)$, $F_2(s)$, $F_3(s)$ are bounded non-decreasing functions on $0 \leqq s \leqq t < 1$, $0 \leqq s \leqq t < 1$, $0 < s < 1$, respectively. This direct analogue of the theorem on non-negative harmonic functions fails for solutions of (1); that is, it is not true that if $u(x,t)$ is a continuous solution of (1) on S and satisfies (11), then $u(x,t)$ is given by (9) in terms of (12_1)-(12_3), where $F_1(s)$, $F_2(s)$, $F_3(s)$ are bounded non-decreasing functions on $0 \leqq s \leqq t < 1$, $0 \leqq s \leqq t < 1$, $0 \leqq s \leqq 1$, respectively. This will follow from (VII) and (XI) below. On the other hand, a certain analogue happens to be true; cf. (VIII) below.

Let $F_1(s)$, $F_2(s)$ be defined for $0 \leqq s < 1$ and be of bounded variation on every fixed interval $0 \leqq s \leqq t < 1$,

$$(13_j)\qquad \int_0^t |\,dF_j(s)\,| < \infty \text{ for } 0 < t < 1 \text{ and } j = 1, 2,$$

and let $F_3(s)$ be defined for $0 < s < 1$ and of bounded variation on every fixed interval $0 < x_0 \leqq s \leqq x_1 < 1$; finally, let

$$(13_3)\qquad \int_0^1 s(1-s)\,|\,dF_3(s)\,| < \infty.$$

The correct analogues of the theorems on non-negative harmonic functions are given by the following theorems:

(VII) *Let* $F_1(s)$, $F_2(s)$, $F_3(s)$ *be non-decreasing functions on* $0 \leqq s < 1$, $0 \leqq s < 1$, $0 < s < 1$, *respectively, satisfying* (13_1)-(13_3). *Then the sum* (9) *of the integrals* (12_1)-(12_3) *represents a non-negative continuous solution of* (1) *on* S.

(VIII) *Conversely, if* $u(x,t)$ *is a non-negative continuous solution of* (1) *on* S, *then there exist non-decreasing functions* $F_1(s)$, $F_2(s)$, $F_3(s)$ *on* $0 \leqq s < 1$, $0 \leqq s < 1$, $0 < s < 1$, *respectively, which satisfy* (13_1)-(13_3), *and* $u(x,t)$ *is given by* (9) *in terms of the functions* (12_1)-(12_3).

It will be clear from the proof that the integrals in (12_1)-(12_3) are ordinary Riemann-Stieltjes integrals (if $\theta_x(x,0)$ is defined to be 0, cf. (17)), and that the integral (12_3) is an absolutely convergent improper Riemann-Stieltjes integral.

There is also an analogue for the theorem of Ostrowski [6] on harmonic functions. This theorem states that if $u(x,t)$ is a continuous solution of (3) on C, then $u(x,t)$ can be represented as the Stieltjes integral of the Poisson kernel with respect to a function of bounded variation if and only if

$$\limsup_{r \to 1-0} \int_0^{2\pi} | u(r\cos\theta, r\sin\theta) | \, d\theta < \infty .$$

For a corresponding theorem for solutions of (2), one might expect that the Poisson-Stieltjes integral should be replaced by the sum of (12_1)-(12_3), where $F_1(s), F_2(s), F_3(s)$ are of bounded variation on $0 \leqq s \leqq t < 1$, $0 \leqq s \leqq t < 1$ and $0 \leqq s \leqq 1$, respectively; and that the condition in the last formula line should be replaced by

$$(14_1) \quad \limsup_{x \to +0} \int_0^t | u(x,s) | \, ds < \infty ; \qquad (14_2) \quad \limsup_{x \to 1-0} \int_0^t | u(x,s) | \, ds < \infty$$

for $0 < t < 1$, and by

$$(14_3{}^*) \qquad \limsup_{t \to +0} \int_0^1 | u(s,t) | \, ds < \infty .$$

Actually, this direct analogue of the theorem of Ostrowski is false; cf. the paragraph following (XI) and its Corollary. In the correct analogue, the true conditions are (14_1)-(14_3), where

$$(14_3) \qquad \limsup_{t\to+0} \int_0^1 s(1-s)\,|\,u(s,t)\,|\,ds < \infty$$

must replace $(14_3{}^*)$, and the more general conditions (13_1)-(13_3) must replace the conditions that $F_1(s)$, $F_2(s)$, $F_3(s)$ are of bounded variation on the respective domains of integration in (12_1), (12_2), (12_3).

The situation is as follows: If $u(x,t)$ is a continuous solution of (1) on S and satisfies (14_1)-(14_2) and $(14_3{}^*)$, then $u(x,t)$ has a representation of the form (9) in terms of (12_1)-(12_3), where $F_1(s)$, $F_2(s)$, $F_3(s)$ are of bounded variation on the corresponding domains of integration. This will be clear from the proof of (X). However, the converse is false; that is, if $u(x,t)$ is given by (9) in terms of (12_1)-(12_3), where $F_1(s)$, $F_2(s)$, $F_3(s)$ are of bounded variation on the domains of integration in (12_1)-(12_3), then (14_1)-(14_3) hold, but $(14_3{}^*)$ can fail to hold. As will be proved below in the paragraph following the italicized assertion (XI) and its Corollary, an example to this effect is furnished by the function (10).

(IX) *Let* $F_1(s)$, $F_2(s)$, $F_3(s)$ *be functions defined on* $0 \leqq s < 1$, $0 \leqq s < 1$, $0 < s < 1$, *respectively, and satisfying* (13_1)-(13_3). *Then the sum* (9) *of the integrals* (12_1)-(12_3) *represents a continuous solution of* (1) *on* S *and satisfies* (14_1)-(14_3).

(X) *Conversely, if* $u(x,t)$ *is a continuous solution of* (1) *on* S *and satisfies* (14_1)-(14_3), *then there exist functions* $F_1(s)$, $F_2(s)$, $F_3(s)$ *defined on* $0 \leqq s < 1$, $0 \leqq s < 1$, $0 < s < 1$, *respectively, satisfying* (13_1)-(13_3), *and* $u(x,t)$ *is given by* (9) *in terms of the functions* (12_1)-(12_3).

As in the case of harmonic functions, there is a uniqueness theorem corresponding to the existence theorems (VII)-(VIII), (IX)-(X). This uniqueness theorem will be deduced from inversion formulae.

(XI) *Let* $F_j(s)$, *where* $j = 1, 2, 3$, *satisfy the conditions of* (IX) *and let* $u(x,t)$ *denote the sum* (9) *of the functions* (12_1)-(12_3). *Then, for* $0 < t < 1$,

$$(15_1) \qquad \lim_{x\to+0} \int_0^t u(x,s)\,ds = F_1(t-0) - F_1(0),$$

$$(15_2) \qquad \lim_{x\to1+0} \int_0^t u(x,s)\,ds = F_2(t-0) - F_2(0)$$

and, for $0 < x < y < 1$,

$$(15_3) \qquad \lim_{t \to +0} \int_x^y u(x,s)\,ds = \tfrac{1}{2}\{F_3(y+0) + F_3(y-0)\} - \tfrac{1}{2}\{F_3(x+0) + F_3(x-0)\}$$

(*the integrals in* (15_1)-(15_3) *exist as Lebesgue integrals*).

These inversion formulae imply the following

Corollary. *If* $u(x,t)$ *is a continuous solution of* (1) *on* S *satisfying* (14_1)-(14_3), *then, in its representation* (9) *in terms of* (12_1)-(12_3), *the functions* $F_1(s)$, $F_2(s)$, $F_3(s)$ *are uniquely determined, up to additive constants, on their continuity points of the open interval* $0 < s < 1$; *also, the jumps* $F_1(+0) - F_1(0)$, $F_2(+0) - F_2(0)$ *are uniquely determined.*

It will now be shown that the function $u = \phi(x,t)$ in (10) has a representation (9) in terms of (12_1)-(12_3), where $F_1(s)$, $F_2(s)$, $F_3(s)$ are of bounded variation on $0 \leqq s \leqq 1$, although $(14_3{}^*)$ is false. The integral in (14_1) and/or (14_2) can be written as the integral of $\exp(-s^2)$ from $s = \frac{1}{2}x/t^{\frac{1}{2}}$ to $s = \infty$ if $\frac{1}{2}x/s^{\frac{1}{2}}$ is introduced as the new integration variable; so that (14_1)-(14_2) hold for $0 < t < 1$. The corresponding integral in (14_3) is, up to a constant factor,

$$\int_0^1 s^2(1-s)t^{-3/2}\exp(-s^2/4t)\,ds < \int_0^1 s^2 t^{-3/2}\exp(-s^2/4t)\,ds.$$

If $\frac{1}{2}s/t^{\frac{1}{2}}$ is introduced as a new integration variable, the last integral becomes the integral of $4s^2\exp(-s^2)$ from $s = 0$ to $s = \frac{1}{2}t^{-\frac{1}{2}}$, which is majorized by

$$4\int_0^\infty s^2\exp(-s^2)\,ds < \infty;$$

so that (14_3) holds. Consequently, (X) implies that (10) has a representation on S of the form (9) in terms of (12_1)-(12_3), where $F_1(s)$, $F_2(s)$, $F_3(s)$ satisfy (13_1)-(13_3). But the inversion formula (15_3) shows that $F_3(s) \equiv \text{const.}$; in fact, the integral in (15_3) is a constant multiple of

$$\int_x^y st^{-3/2}\exp(-s^2/4t)\,ds = 2t^{-\frac{1}{2}}\{\exp(-x^2/4t) - \exp(-y^2/4t)\},$$

which tends to 0, as $t \to 0$, for $0 < x < y$. Thus, not only do the functions

$F_1(s)$, $F_2(s)$, $F_3(s)$ satisfy (13_1)-(13_3) but they are of bounded variation, on $0 \leqq s \leqq 1$, as well. Nevertheless, $(14_3{}^*)$ fails to hold, since, to a constant factor, the integral in $(14_3{}^*)$ is

$$\int_0^1 st^{-3/2} \exp(-s^2/4t)\, ds = 2t^{-\frac{1}{2}}\{1 - \exp(-1/4t)\}$$

and therefore tends to ∞, as $t \to +0$.

It may be mentioned that in the representation of the function (10) as the sum of (12_1)-(12_3), the functions F_1, F_2, F_3, up to additive constants, are given by

$$F_1(0) = 0, \qquad F_1(s) = \tfrac{1}{2}\pi^{-\frac{1}{2}} \text{ if } s > 0;$$

$$F_2(s) = \tfrac{1}{2}\pi^{-\frac{1}{2}} \int_0^s r^{-3/2} \exp(-1/4r)\, dr; \qquad F_3(s) \equiv 0.$$

Still another theorem, analogous to one for harmonic functions, deals with the existence of one-dimensional boundary limits:

(XII) *Let* $F_1(s)$, $F_2(s)$, $F_3(s)$ *satisfy the conditions* of (IX) *and let* $u(x,t)$ *denote the sum* (9) *of the functions* (12_1)-(12_3). *Then, for* $j = 1, 2, 3$, *the limit* (6_j) *exists at every* s *on* $0 < s < 1$ *at which the derivative* $F_j'(s)$ *exists; in fact, for such an s-value,* $f_j(s) = F_j'(s)$. (*In particular, the limits* (6_1)-(6_3) *exist almost everywhere.*)

It may be remarked that if $j = 1$, 2 or 3, then (12_j) reduces to (8_j) if and only if $F_j(s)$ is absolutely continuous. Consequently, a solution $u(x,t)$ of (1) on S can possess limits (6_1)-(6_3), which are bounded, without $u(x,t)$ being bounded on S and, therefore, without possessing a representation as the sum of integrals (8_1)-(8_3). In fact, if $u(x,t)$ is representable as the sum of integrals (12_1)-(12_3), then $u(x,t)$ is bounded on S if and only if $F_1(s)$, $F_2(s)$, $F_3(s)$ are absolutely continuous on the domains of integration and possess bounded derivatives almost everywhere.

Before proving the assertions (V)-(XII), it will be convenient first to prove a part of (IX).

First part of the proof of (IX). It is readily verified from the identity

$$\theta(x,t) = \pi^{-\frac{1}{2}} \sum_{k=-\infty}^{+\infty} t^{-\frac{1}{2}} \exp(-(x+2k)^2/4t), \qquad t > 0, \tag{16}$$

supplied by the linear transformation of the ϑ_3-function, that

$$\theta_x(x,t) \to 0 \text{ as } t \to +0, \qquad (0 < x < 2). \tag{17}$$

In fact, the series (16) and its derived series are uniformly convergent for $0 < x_0 \leqq x \leqq x_1 < 2$ and $0 < t \leqq t_1 < \infty$; while, on this x-range, each term tends uniformly to 0 as $t \to +0$ (cf. the formulae following (53_1) below). Hence, (17) holds uniformly on such an x-interval. Clearly, (17) implies that, if $F_1(s)$, $F_2(s)$ are of bounded variation on $0 \leqq s \leqq t$, then the integrals in (12_1), (12_2) exist, as Riemann-Stieltjes integrals, for every fixed (x, t) in S. Furthermore, the remark concerning the uniformity of (17) shows that (12_1), (12_2) are continuous functions on S.

That (12_1) is a solution of (1) on S can be proved directly. However, in order to simplify matters in subsequent proofs, a somewhat indirect proof will be employed. Put

$$U_1(x, t) = \int_0^t u_1(x, r)\,dr. \tag{18}$$

Since, formally,

$$\int_0^t u_1(x, r)\,dr = -\int_0^t \int_0^r \theta_x(x, r-s)\,dF_1(s))\,dr,$$

the existence of the integral in (18) will be proved if it is shown that the iterated integral

$$-\int_0^t \int_s^t \theta_x(x, r-s)\,dr\,dF_1(s) = -\int_0^t \int_0^{t-s} \theta_x(x, r)\,dr\,dF_1(s)$$

converges absolutely. To this end, since $F_1(s)$ is of bounded variation on $0 \leqq s \leqq t$, it can be supposed that $F_1(s)$ is non-decreasing (otherwise $F_1(s)$ is written as the difference of two non-decreasing functions, and each of the corresponding integrals is treated separately). Also, since

$$-\theta_x(x, t) > 0 \text{ for } 0 < x < 1 \text{ and } 0 < t < \infty \tag{19}$$

(cf., e. g., [8], p. 410), the last iterated integral can be treated without the insertion of absolute value signs. But (17) and an integration by parts show that the integral exists and is

$$U_1(x, t) = -\int_0^t \theta_x(x, t-s)\,(F_1(s) - F_1(0))\,ds. \tag{20}$$

Thus, (18) can be written in the form (8_1), where $f_1(s) = F_1(s) - F_1(0)$ is Riemann integrable for $0 \leqq s \leqq t < 1$ (for t fixed). Consequently, by standard theorems (cf., e. g., [1], p. 355), (18) is a continuous solution of (1) on S. Hence, by (I), the function (18) is of class C^∞ on S. Since (12_1) is continuous on S, the function $u_1(x,t) = \partial U_1(x,t)/\partial t$ is a continuous solution of (1) on S.

Similarly,

$$U_2(x,t) = \int_0^t u_2(x,r)\,dr \tag{21}$$

exists as a Lebesgue integral. Furthermore,

$$U_2(x,t) = \int_0^t \theta_x(x-1,t-s)(F_2(s)-F_2(0))\,ds \tag{22}$$

is of class C^∞ and a solution of (1) on S. Hence, the function (12_2), which is $\partial U_2/\partial t$, is a continuous solution of (1) on S.

In order to see that (12_3) exists and is a continuous solution of (1) on S, note that (7) implies that

$$\tfrac{1}{2}\{\theta(x-s,t)-\theta(x+s,t)\} = 2\sum_{k=1}^{\infty}\exp(-k^2\pi^2 t)\sin k\pi x \sin k\pi s. \tag{23}$$

Hence, for a fixed (x,t) in S, this function vanishes as s (or $1-s$) at $s=0$ (or $s=1$). Consequently, if $F_3(s)$ is defined on $0<s<1$ and satisfies (13_3), the integral in (12_3) converges absolutely as an improper Riemann-Stieltjes integral. In fact, since (13_3) implies that

$$\int_0^1 |\sin k\pi s|\,|dF_3(s)| \text{ exists and is } O(k), \text{ as } k\rightarrow\infty, \tag{24}$$

the series (23) can be integrated term-by-term. Clearly,

$$u_3(x,t) = 2\sum_{k=1}^{\infty}\exp(-k^2\pi^2 t)\sin k\pi x \int_0^1 \sin \pi s\, dF_3(s) \tag{25}$$

is defined and continuous on S. Since $\exp(-k^2\pi^2 t)\sin k\pi x$, where $k=1, 2,\cdots$, is a solution of (1), and since (25) can be differentiated (repeatedly) term-by-term at any point (x,t) of S, it follows that (25), that is, (12_3), is a continuous solution of (1) on S.

In order to complete the proof of (IX), it remains to verify that the sum (9) of (12_1)-(12_3) satisfies (14_1)-(14_3). This will be done below with the aid of Lemma 1.

Proof of (VII). In view of the portion of (IX) just proved, it only remains to verify that the function (12_j) is non-negative when $dF_j \geqq 0$. For $j = 1$, this is clear from (19). Since, for a fixed t, the function $\theta_x(x, t)$ is even and of period 2 in x, it is seen from (19) that

$$\theta_x(x-1, t) > 0 \text{ for } 0 < x < 1 \text{ and } 0 < t < \infty. \tag{26}$$

Hence (12_2) is non-negative. Finally, the kernel (23) of (12_3) satisfies

$$\theta(x-s, t) - \theta(x+s, t) > 0 \text{ for } 0 < x < 1,\ 0 < s < 1,\ t > 0, \tag{27}$$

by (19); so that (12_3) is non-negative.

In order to complete the proof of (IX), a lemma will first be established.

LEMMA 1. *Let* $u(x, t)$ *be a non-negative, continuous solution of* (1) *on* S. *Then, if* $0 < x < 1$, *the integrals occurring in* (14_1)-(14_2) *exist as Lebesgue integrals and, if* $0 < t < 1$, *the integral in* (14_3) *exists as a Lebesgue integral. Furthermore,* (14_1)-(14_3) *hold.*

Proof of Lemma 1. Let ϵ, ξ, τ denote numbers satisfying

$$\tfrac{1}{2} > \epsilon > \xi > 0 \text{ and } \tfrac{1}{2} > \epsilon > \tau > 0. \tag{28}$$

Since $u(x, t)$ is a continuous solution of (1) on S, the function $u = v(x, t) = v(x, t, \epsilon, \xi, \tau)$ defined by

$$v(x, t) = u(x(1-\epsilon) + \xi,\ t(1-\epsilon)^2 + \tau) \tag{29}$$

is a solution of (1) on S and is uniformly continuous on the open set S. Hence, by (II) and (IV), the function $u = v(x, t)$ has a representation of the form (9) in terms of (8_1)-(8_3), where the functions $f_j(s) = f_j(s, \epsilon, \xi, \tau)$; $j = 1, 2, 3$, are given by

$$f_1(s) = u(\xi, s(1-\epsilon)^2 + \tau); \tag{30_1}$$

$$f_2(s) = u(1-\epsilon+\xi, s(1-\epsilon)^2 + \tau) \tag{30_2}$$

and

$$f_3(s) = u(s(1-\epsilon) + \xi, \tau). \tag{30_3}$$

Since $u(x, t)$ is non-negative, the inequalities (19), (26) and (27) show that each of the corresponding terms of (9) is non-negative and, therefore, none of them exceeds their sum, $v(x, t)$. Hence, if $0 < \delta < t$,

$$v(x, t) \geqq -\int_0^t \theta_x(x, t-s) f_1(s)\, ds \geqq -\int_0^{t-\delta} \theta_x(x, t-s) f_1(s)\, ds. \tag{31}$$

If (x, t) is an arbitrary point of R, there exists a constant $C = C(x, t, \delta)$ satisfying

$$-\theta_x(x, t-s) \geqq C > 0 \text{ for } 0 \leqq s \leqq t-\delta. \tag{32}$$

This can be deduced, for instance, from the Jacobi identity

$$\theta(x, t) = \prod_{n=1}^{\infty} (1-q^{2n})(1+2q^{2n-1}\cos \pi x + q^{4n-2}),$$

where $q = \exp(-\pi^2 t)$. This identity shows the existence of a constant, depending on δ, for which $\theta(x, t) \geqq \text{const.} > 0$ if $0 < \delta < t \leqq 1$. Also, logarithmic differentiation of the Jacobi identity with respect to x gives

$$-\theta_x(x, t)/\theta(x, t) = 2\pi \sin \pi x \sum_{n=1}^{\infty} q^{2n-1}(1+2q^{2n-1}\cos \pi x + q^{4n-2})^{-1}.$$

Clearly, (31) follows from the last two formula lines in view of the definition of q. (A similar use of the Jacobi identity was made in [8], p. 410.) But (31) and (32) imply that

$$C\int_0^{t-\delta} f_1(s)\,ds \leqq v(x, t).$$

In view of the definitions (30_1) of $f_1(s)$ and (29) of $v(x, t)$, this inequality can be written as

$$C\int_0^{t-\delta} u(\xi, s(1-\epsilon)^2+\tau)\,ds \leqq u(x(1-\epsilon)+\xi, t(1-\epsilon)^2+\tau).$$

If δ, ϵ, ξ are fixed, but $\tau \rightarrow 0$ (which is permissible in view of the second set of inequalities in (28)), it follows from Fatou's lemma that

$$C\int_0^{t-\delta} u(\xi, s(1-\epsilon)^2)\,ds \leqq u(x(1-\epsilon)+\xi, t(1-\epsilon)^2),$$

or, by a change of variables,

$$C(1-\epsilon)^{-2}\int_0^{t} u(\xi, s)\,ds \leqq u(x(1-\epsilon)+\xi, t+\delta(1-\epsilon)^2),$$

where t now stands for $(t-\delta)(1-\epsilon)^2$.

Since ϵ, δ (> 0) are arbitrary, it is seen that the integral in (14_1) and/or (14_2) exists. If δ, ϵ (> 0) are kept fixed, while $\xi \rightarrow 0$, the last formula line implies

$$\limsup_{\xi\to 0}\int_0^t u(\xi, s)\,ds \leqq (1-\epsilon)^2 u(x(1-\epsilon), t+\delta(1-\epsilon)^2)/C < \infty,$$

that is, that (14_1) holds. In a similar manner, it can be shown that (14_2) holds.

The proof of the existence of the integral in (14_3) requires a slight modification of the procedure above. By the argument leading to (31),

$$v(x,t) \geqq \tfrac{1}{2}\int_0^1 \{\theta(x-s,t)-\theta(x+s,t)\}f_3(s)\,ds. \tag{33}$$

For a fixed $t > 0$, the function $\theta(x,t)$ is even, of period 2 in x, and strictly increasing for $0 < x < 1$ (cf. (19)). Hence to any $\delta > 0$, there corresponds a constant $C = C(\delta)$ such that

$$\tfrac{1}{2}\{\theta(x-s,t)-\theta(x-s,t)\} \geqq C > 0 \text{ for } \delta \leqq s \leqq 1-\delta.$$

In addition,

$$\tfrac{1}{2}\{\theta(x-s,t)-\theta(x+s,t)\}/s \to -\theta_x(x,t) \text{ as } s\to 0 \tag{34}$$

and

$$\tfrac{1}{2}\{\theta(x-s,t)-\theta(x-s,t)\}/(1-s) \to \theta_x(x+1,t) \text{ as } s\to 1. \tag{35}$$

Consequently,

$$\tfrac{1}{2}\{\theta(x-s,t)-\theta(x+s,t)\} \geqq C^* s(1-s) \text{ for } 0\leqq s\leqq 1$$

for some positive constant C^*, depending on the point (x,t) of S. Thus, (33) shows that

$$C^*\int_0^1 s(1-s)f_3(s)\,ds \leqq v(x,t).$$

In view of the definitions, (30_3) of $f_3(s)$ and (29) of $v(x,t)$,

$$C^*\int_0^1 s(1-s)u(s(1-\epsilon)+\xi,\tau)\,ds \leqq u(x(1-\epsilon)+\xi, t(1-\epsilon)^2+\tau).$$

Keeping ϵ, τ fixed, let $\xi\to 0$. Then, by Fatou's lemma,

$$C^*\int_0^1 s(1-s)u((s(1-\epsilon),\tau)\,ds \leqq u(x(1-\epsilon, t(1-\epsilon)^2+\tau).$$

A change of variables shows that, if $\epsilon > 0$ is sufficiently small,

$$C^*(1-\epsilon)^{-2}\int_0^{\frac{1}{2}} s(1-s)u(s,\tau)\,ds \leqq u(x(1-\epsilon), t(1-\epsilon)^2+\tau).$$

A repetition of the above arguments, with x replaced by $1-x$, leads to

$$C^{**}(1-\epsilon)^{-2}\int_{\frac{1}{2}}^{1} s(1-s)u(s,\tau)\,ds \leqq u((1-x)(1-\epsilon), t(1-\epsilon)^2+\tau),$$

where C^{**} is a positive constant depending on the point $(1-x, t)$ of S.

These inequalities imply that the integral in (14_3) exists as a Lebesgue integral. By keeping $\epsilon > 0$ fixed and letting $\tau \rightarrow 0$ in the last two formula lines, the upper limit relation (14_3) follows. This completes the proof of Lemma 1.

Completion of the proof of (IX). It is sufficient to verify that each of the functions $u = u_j(x,t)$ given by (12_j), where $j = 1, 2, 3$, satisfies (14_1)-(14_3). For, if (14_1)-(14_3) holds for two or more functions, it clearly holds for their sum. In (12_1), it can be supposed that $F_1(s)$, a function of bounded variation in $0 \leqq s \leqq t < 1$, is non-decreasing (for otherwise $F_1(s)$ is written as the difference of two non-decreasing functions, and each of the two corresponding functions (12_1) is treated separately). But if $F_1(s)$ is non-decreasing, then (VII), proved above, implies that (12_1) is non-negative. Then (14_1)-(14_3) follow from Lemma 1. In a similar manner, it is shown that $u = u_2(x,t)$, given by (12_2), satisfies (14_1)-(14_3).

If a function $F_3(s)$, defined on $0 < s < 1$ and satisfying (13_3), can be written as the difference of two non-decreasing functions each of which satisfies (13_3), then it follows, in the same way as above, that $u = u_3(x,t)$, given by (12_3), satisfies (14_1)-(14_3). Finally, it is clear that the usual device used for decomposing functions of bounded variation,

$$F_3(s) = \int_{\frac{1}{2}}^{s} |\,dF_3(s)\,| - \left(\int_{\frac{1}{2}}^{s} |\,dF_3(s)\,| - F_3(s)\right), \qquad (0 < s < 1),$$

yields the desired decomposition. This completes the proof of (IX).

Proof of (VIII). It will be shown that (VIII) follows from Lemma 1 and from the assertions (X) and (XI), to be proved below. In fact, Lemma 1 and (X) imply that, if $u(x,t)$ is a non-negative, continuous solution of (1) on S, then there exist functions $F_1(s)$, $F_2(s)$, $F_3(s)$ satisfying (13_1)-(13_3)

and having the property that $u(x, t)$ possesses a representation (9) in terms of (12_1)-(12_3). Finally, the inversion formulae in (XI) and the non-negativity of $u(x, t)$ imply that the functions $F_1(s)$, $F_2(s)$, $F_3(s)$ are non-decreasing.

Proof of (X). Let $u(x, t)$ be a continuous solution of (1) on S satisfying (14_1)-(14_3). Let $0 < \epsilon < \frac{1}{2}$ and put

$$u^\epsilon(x, t) = u(x(1-2\epsilon) + \epsilon, t(1-2\epsilon)^2 + \epsilon). \tag{36}$$

Then $u^\epsilon(x, t)$ is uniformly continuous on S and represents a solution of (1) on S. By (II) and (IV), the function $u = u^\epsilon(x, t)$ possesses a representation (9) in terms of functions (8_1)-(8_3). The Riemann integrals (8_1)-(8_3) can be written as Riemann-Stieltjes integrals (12_1)-(12_3). For the case at hand, the function $F_1(s) = F_1{}^\epsilon(s)$ is given by

$$F_1{}^\epsilon(s) = \text{const.} + \int_0^s u(\epsilon, r(1-2\epsilon)^2 + \epsilon)\,dr,$$

or, for a suitable choice of the const., by

$$F_1{}^\epsilon(s) = (1-2\epsilon)^{-2} \int_0^{s(1-2\epsilon)^2+\epsilon} u(\epsilon, r)\,dr, \qquad (0 < s < 1). \tag{37_1}$$

Similarly,

$$F_2{}^\epsilon(s) = (1-2\epsilon)^{-2} \int_0^{s(1-2\epsilon)^2+\epsilon} u(1-\epsilon, r)\,dr, \qquad (0 < s < 1) \tag{37_2}$$

and

$$F_3{}^\epsilon(s) = (1-2\epsilon)^{-1} \int_{\frac{1}{2}}^{s(1-2\epsilon)+\epsilon} u(r, \epsilon)\,dr, \qquad (0 < s < 1). \tag{37_3}$$

Hence

$$u^\epsilon(x, t) = u_1{}^\epsilon(x, t) + u_2{}^\epsilon(x, t) + u_3{}^\epsilon(x, t), \tag{9^ϵ}$$

where

$$u_1{}^\epsilon(x, t) = -\int_0^t \theta_x(x, t-s)\,dF_1{}^\epsilon(s), \tag{$12_1{}^\epsilon$}$$

$$u_2{}^\epsilon(x, t) = \int_0^t \theta_x(x-1, t-s)\,dF_2{}^\epsilon(s), \tag{$12_2{}^\epsilon$}$$

$$(12_3{}^\epsilon) \qquad u_3{}^\epsilon(x,t) = \tfrac{1}{2}\int_0^1 \{\theta_x(x-s,t) - \theta_x(x+s,t)\} dF_3{}^\epsilon(s).$$

Clearly, the assumptions (14_1)-(14_3) imply the existence of the functions (37_1)-(37_3). Also, (14_1) implies that, for any t on $0 < t < 1$, the functions $F_1{}^\epsilon(s)$ are uniformly bounded and of uniform bounded variation for $0 \leqq s \leqq t$. Helly's selection theorem implies therefore the existence of a function $F_1(s)$, $0 \leqq s < 1$, which is of bounded variation on every interval $0 \leqq s \leqq t < 1$, and the existence of a sequence of positive numbers $\epsilon_1{}^1, \epsilon_2{}^1, \cdots$ satisfying

$$\lim_{n\to\infty} \epsilon_n{}^1 = 0 \text{ and } \lim_{n\to\infty} F_1{}^\epsilon(s) = F_1(s), \qquad (\epsilon = \epsilon_n{}^1),$$

where $0 < s < 1$. In addition, (17) and Helly's term-by-term integration for Riemann-Stieltjes integrals imply that the function $(12_1{}^\epsilon)$ where $\epsilon = \epsilon_n{}^1$, tends to (12_1) as $n \to \infty$, for every (x, t) in S.

The arguments applied to the family of functions $F_2{}^\epsilon(s)$, where $\epsilon = \epsilon_n{}^1$, also show the existence of a function $F_2(s)$, of bounded variation on every interval $0 \leqq s \leqq t < 1$, and of a subsequence $\epsilon_1{}^2, \epsilon_2{}^2, \cdots$ of the sequence $\epsilon_1{}^1. \epsilon_2{}^1, \cdots$, having the property that $(12_2{}^\epsilon)$, where $\epsilon = \epsilon_n{}^2$, tends to (12_2) as $n \to \infty$, for every (x, t) in S.

Finally, by considering the family of functions $F_3{}^\epsilon(s)$, where $\epsilon = \epsilon_n{}^2$, it is seen that there exist a function $F_3(s)$, $0 < s < 1$, of bounded variation on $\delta \leqq s \leqq 1 - \delta$ for every small positive δ, and a subsequence $\epsilon_1{}^3, \epsilon_2{}^3, \cdots$ of the sequence $\epsilon_1{}^2, \epsilon_2{}^2, \cdots$ in such a way that

$$\lim_{n\to\infty} F_3{}^\epsilon(s) = F_3(s) \qquad (\epsilon = \epsilon_n{}^3)$$

holds on $0 < s < 1$. Furthermore, (14_3) implies that

$$\limsup_{\epsilon\to 0} \int_0^1 s(1-s)\,|\,dF_3{}^\epsilon(s)\,| < \infty, \text{ hence} \int_0^1 s(1-s)\,|\,dF_3(s)\,| < \infty.$$

Hence, if $G^\epsilon(s)$ and $G(s)$ are defined by $dG^\epsilon(s) = s(1-s)dF_3{}^\epsilon(s)$ and $dG(s) = s(1-s)dF_3(s)$, respectively, these functions are of uniformly bounded variation. If the kernel in the integral in $(12_3{}^\epsilon)$ is divided by $s(1-s)$, the relations (34) and (35) and an application of Helly's term-by-term integration theorem show that $(12_3{}^\epsilon)$ tends to (12_3) as $n \to \infty$ (where $\epsilon = \epsilon_n{}^3$).

Consequently, if $\epsilon = \epsilon_n{}^3$ in (9^ϵ) and $n \to \infty$, the right side of (9^ϵ) tends to (9), where u_1, u_2, u_3 are given by (12_1), (12_2), (12_3), respectively. On the other hand, the left side of (9^ϵ) tends to $u(x, t)$ as $\epsilon \to 0$, by virtue of

(36) and the continuity of $u(x,t)$ on S. This completes the proof of the representation (9) for the given function $u(x,t)$ in terms of integrals of the form (12_1)-(12_3).

For the proof of (XI) and (XII), it will be convenient to have the following localization theorem:

LEMMA 2. *Let $F_3(s)$ be defined on $0 < s < 1$, be of bounded variation on every closed interval of $0 < s < 1$, and such as to satisfy (13_3). Let $0 < \delta < 1$ and $0 < \delta < x < 1 - \delta$. Then*

$$(38) \qquad \lim_{t \to +0} \tfrac{1}{2} \int_0^\delta \{\theta(x-s,t) - \theta(x+s,t)\} dF_3(s) = 0$$

and

$$(39) \qquad \lim_{t \to +0} \tfrac{1}{2} \int_{1-\delta}^1 \{\theta(x-s,t) - \theta(x+s,t)\} dF_3(s) = 0,$$

and both of these limit relations are uniform with respect to x and δ if $0 < \text{const.} \leqq x - \delta$ and $1 - x - \delta \geqq \text{const.} > 0$.

Proof of Lemma 2. If $s > 0$, the mean-value theorem of differential calculus gives

$$\tfrac{1}{2}\{\theta(x-s,t) - \theta(x+s,t)\}/s = -\theta_x(x+\xi s, t),$$

where $|\xi| < 1$ and $\xi = \xi(x,s,t)$. Since (17) holds uniformly on any interval $0 < x_0 \leqq x \leqq x_1 < 2$, it follows that, if $\epsilon > 0$ is given,

$$\tfrac{1}{2}\,|\,\theta(x-s,t) - \theta(x+s,t)\,|/s < \epsilon$$

holds for $0 < s \leqq \delta$ and for sufficiently small $t > 0$. But then the integral in (38) is majorized by $\epsilon \int_0^\delta s\,|\,dF_3(s)|$ for small $t > 0$. In view of (13_3), this proves (38). The relation (39) is proved in the same way; cf. (35).

The last part of Lemma 2, concerning the uniformity of (38) and (39), is clear from the above proof.

LEMMA 3. *Let $F_3(s)$ be defined on $0 < s < 1$, be of bounded variation on every closed interval on $0 < s < 1$, and such as to satisfy (13_3). Then the series*

$$(40) \qquad 2 \sum_{k=1}^{\infty} (k\pi)^{-2} \sin k\pi x \int_0^1 \sin k\pi s \, dF_3(s) = G(x)$$

is uniformly convergent for $0 \leqq x \leqq 1$; *in particular*

$$\lim_{x\to 0} G(x) = 0 \text{ and } \lim_{x\to 1} G(x) = 0. \tag{41}$$

Proof. First, it will be shown that (13_3) implies

$$s\int_s^{\frac{1}{2}} |\, dF_3(r)\,| \to 0 \text{ and } s\int_{\frac{1}{2}}^{1-s} |\, dF_3(r)\,| \to 0, \qquad (s\to 0), \tag{42}$$

and

$$\int_0^{\frac{1}{2}}\int_s^{\frac{1}{2}} |\, dF_3(r)\,|\, ds < \infty \text{ and } \int_{\frac{1}{2}}^{1}\int_{\frac{1}{2}}^{s} |\, dF_3(r)\,|\, ds < \infty. \tag{43}$$

Let $\epsilon > 0$ be given and let $\delta = \delta_\epsilon > 0$ be chosen so small that

$$\epsilon > \int_0^{\delta} r\,|\, dF_3(r)\,| \geqq \int_s^{\delta} r\,|\, dF_3(r)\,| \geqq s\int_s^{\delta} |\, dF_3(r)\,|\,.$$

Since, if $\delta > 0$ fixed, $s\int_{\delta}^{\frac{1}{2}} |\, dF_3(r)\,| \to 0$ as $s \to 0$, the first part of (42) follows, $\epsilon > 0$ being arbitrary. The second part of (42) is proved in the same way.

By Fubini's theorem on iterated integrals, the integrals in (43) are

$$\int_0^{\frac{1}{2}} s\,|\, dF_3(s)\,| \text{ and } \int_{\frac{1}{2}}^{1} (1-s)\,|\, dF_3(s)\,|,$$

respectively, and are finite by (13_3).

Let $G(x)$ denote, for a moment, the function defined on $0 \leqq x < \frac{1}{2}$ by

$$G(x) = \int_0^{x}\int_s^{\frac{1}{2}} dF_3(r)\, ds + \alpha x \tag{44_1}$$

and on $\frac{1}{2} < x \leqq 1$ by

$$G(x) = \int_x^{1}\int_{\frac{1}{2}}^{s} dF_3(r)\, ds + \alpha x + \beta. \tag{44_2}$$

In view of (43), these definitions are meaningful. If the constants α and β are chosen to be

$$\alpha = -\beta = \int_{\frac{1}{2}}^{1}\int_{\frac{1}{2}}^{s} dF_3(r)\, ds - \int_0^{\frac{1}{2}}\int_s^{\frac{1}{2}} dF_3(r)\, ds,$$

it is seen that $G(\frac{1}{2}-0)=G(\frac{1}{2}+0)$; so that $G(x)$ can be defined at $x=\frac{1}{2}$ so as to be continuous there. The above choice of α, β assures that

$$G(0)=G(1)=0. \tag{45}$$

The relations (43) show that $G(x)$ is absolutely continuous for $0\leqq x\leqq 1$ and that

$$G'(x) \text{ is } \int_{x}^{\frac{1}{2}} dF_3(r)+\alpha \text{ or } -\int_{\frac{1}{2}}^{x} dF_3(r)+\alpha \tag{46}$$

according as $0<x\leqq\frac{1}{2}$ or $\frac{1}{2}\leqq x<1$. The k-th Fourier sine coefficient of $G(x)$ on $0<x<1$ is, by (45),

$$2\int_0^{\pi} G(x)\sin k\pi x\,dx=2(k\pi)^{-1}\int_0^{\pi} G'(x)\cos k\pi x\,dx,$$

the last integral existing in virtue of (43). Since (42) implies that $G'(x)\sin k\pi x\to 0$ as $x\to+0$ or $x\to 1-0$, another integration by parts and (46) show that the expression on the right in the last formula line is

$$2(k\pi)^{-2}\int_0^1 \sin k\pi x\,dF_3(x).$$

Hence the series in (40) is the Fourier sine series for $G(x)$. Since $G(x)$ is of bounded variation, continuous on $0\leqq x\leqq 1$, and vanishes at $x=0$ and $x=1$, the uniform convergence of the series (40) follows. This proves Lemma 3.

Proof of (XI). Let $u(x,t)$ be a continuous solution of (1) on S, representable in the form (9) in terms of functions (12_1)-(12_3). It will first be shown that, for $0<t<1$,

$$\lim_{x\to+0}\int_0^t u_1(x,s)\,ds=F_1(t-0)-F_1(0). \tag{47_1}$$

In the proof of (47_1), use will be made of (18) and (20), as follows: The function $U_1(x,t)$ has a representation of the form (9) in terms of functions of the form (8_1)-(8_3), where $f_1(s)$, $f_2(s)$, $f_3(s)$ are $F_1(s)-F(0)$, 0, 0, respectively. Hence, by standard theorems (cf. [1], p. 355), (47_1) holds almost everywhere. Actually, the limit in (6_1) exists and is $f_1(s-0)$ for every point s at which $f_1(s)$ is continuous from the left (cf. [1], p. 355).

(This can be deduced from the standard treatment of Lebesgue for "singular integrals"; cf. the proof of (XII). Incidentally, (47_1) follows from (XII) itself.) Hence, (47_1) holds for every t on $0 < t < 1$.

These arguments also show that

$$(47_2) \qquad \lim_{x\to 1-0} \int_0^t u_1(x, s)\,ds = 0, \qquad (0 < t < 1).$$

The limit relation

$$(47_3) \qquad \lim_{t\to +0} \int_0^t u_1(x, s)\,ds = 0, \qquad (0 < x < 1),$$

is a consequence of the existence of the integral (47_3) as a Lebesgue integral.

In the same way as (18), (20) lead to (47_1) and (47_2), it is seen that (21), (22) lead to

$$(48_1) \qquad \lim_{x\to +0} \int_0^t u_2(x, s)\,ds = 0, \qquad (0 < t < 1),$$

and

$$(48_2) \qquad \lim_{x\to 1-0} \int_0^t u_2(x, s)\,ds = F_2(t-0) - F_2(0) \qquad (0 < t < 1).$$

In addition, the proof of (47_3) shows that

$$(48_3) \qquad \lim_{t\to +0} \int_0^t u_2(x, s)\,ds = 0, \qquad (0 < x < 1).$$

Finally, it will be shown that

$$(49_1) \qquad \lim_{x\to +0} \int_0^t u_3(x, s)\,ds = 0, \qquad (0 < t < 1)$$

and

$$(49_2) \qquad \lim_{x\to 1-0} \int_0^t u_3(x, s)\,ds = 0, \qquad (0 < t < 1)$$

and that, for $0 < x < y < 1$,

$$(49_3) \qquad \lim_{t\to +0} \int_x^y u_3(s, t)\,ds = \tfrac{1}{2}\{F_3(y+0) + F_3(y-0)\} - \tfrac{1}{2}\{F_3(x+0) + F_3(x-0)\}.$$

If $0 < \epsilon < t$, it is clear from (25) that

$$\int_{\epsilon}^{t} u_3(x, s)\, ds = 2 \sum_{k=1}^{\infty} (k\pi)^{-2}\{\exp(-k^2\pi^2\epsilon) - \exp(-k^2\pi^2 t)\} \sin k\pi x \int_0^1 \sin k\pi s\, dF_3(s).$$

Since the integral in (49_1) and/or (49_2) exists, the convergence of (40) and the regularity of Abel's summation method show that the last equation becomes

$$\int_0^t u_3(x, s)\, ds = 2 \sum_{k=1}^{\infty} (k\pi)^{-2}\{1 - \exp((-k^2\pi^2 t)\} \sin k\pi x \int_0^1 \sin k\pi s dF_3(s),$$

as $\epsilon \rightarrow +0$. (Strictly speaking, $\epsilon \rightarrow +0$ is not Abel's summation method; however, if 0 terms are inserted to take care of terms $\exp(-k\pi^2\epsilon)$, for k not a square, Abel's theorem becomes applicable.) For a fixed $t > 0$, the last series is uniformly convergent for $0 \leqq x \leqq 1$. Hence (49_1), (49_2) follow.

In order to prove (49_3), let δ be so chosen that $0 < \delta < x < y < 1 - \delta$. Then the uniformity statement concerning (38), (39) in Lemma 2 implies that the expression on the left of (49_3) is the limit of

$$\int_x^y \tfrac{1}{2} \int_{\delta}^{1-\delta} \{\theta(s - r, t) - \theta(s + r, t)\} dF_3(r)\, ds$$

as $t \rightarrow +0$, provided that this limit exists. For a fixed $t > 0$, the expression in the last formula line is

$$(50) \qquad 2 \sum_{k=1}^{\infty} (k\pi)^{-1} \exp(-k^2\pi^2 t)\{\cos k\pi y - \cos k\pi x\} \int_{\delta}^{1-\delta} \sin k\pi s\, dF_3(s).$$

Consider a function of bounded variation $H(s)$ defined on $0 \leqq s \leqq 1$ by letting $dH(s)$ be 0, $dF(s)$ or 0 according as $0 \leqq s < \delta$, $\delta < s < 1 - \delta$ or $1 - \delta < s \leqq 1$. If the Fourier-Stieltjes sine series of $H(s)$, on $0 \leqq s \leqq 1$, is integrated from $s = x$ to $s = y$, the result is (50), when $t = 0$. Consequently, when $t = 0$, the series (50) is convergent and its sum is the expression on the right of (49_3), since $0 < \delta < x < y < 1 - \delta$. Finally, (49_3) follows, since $t \rightarrow +0$ in (50) is a regular summation process; cf. the remarks in the proof of (49_1)-(49_2).

The inversion formulae (15_1), (15_2), (15_3) are consequences of (9) and (47_1)-(47_3), (48_1)-(48_3), (49_1)-(49_3), respectively. This completes the proof of (XI).

The proof of (XII) will depend on Lemma 2 and a standard lemma of the Lebesgue-Toeplitz type; cf. [4].

LEMMA 4. *For a fixed* τ *on* $0 < \tau \leqq 1$, *let* $\psi(\sigma, \tau)$ *be a continuous function of bounded variation on the interval* $|\sigma| \leqq l \leqq \infty$, *and possess the following properties*:

(i) $\lim\limits_{\tau \to +0} \psi(\sigma, \tau) = 0$ for every fixed σ on $0 < |\sigma| \leqq l$;

(ii) $\lim\limits_{\tau \to +0} \int\limits_{|\sigma| > \delta} |d_\sigma \psi(\sigma, \tau)| = 0$ for every fixed δ on $0 < \delta < l$;

(iii) $\limsup\limits_{\tau \to +0} \int\limits_{|\sigma| < \eta} |\sigma| \, |d_\sigma \psi(\sigma, \tau)| < \infty$ for some $\eta > 0$;

(iv) $\lim\limits_{\tau \to +0} \int\limits_{|\sigma| < \delta} \sigma d_\sigma \psi(\sigma, \tau) = -1$ for every fixed δ on $0 < \delta < \eta$.

Let $F(x)$ *be defined and bounded for* $|x| \leqq l$ *and of bounded variation on every closed interval on* $|x| < l$. *Then*

$$I(x, \tau) = \int_{\alpha}^{\beta} \psi(x - \sigma, \tau) \, dF(\sigma), \tag{51}$$

where $\alpha = \max(-l, x - l)$ *and* $\beta = \min(l, x + l)$, *exists and, for every* x *on* $|x| < l$ *at which* $F(x)$ *is differentiable, satisfies*

$$\lim_{\tau \to +0} I(x, \tau) = F'(x). \tag{52}$$

Proof of Lemma 4. The $dF(\sigma)$ in (51) can be replaced by $d\{F(\sigma) - F(x)\}$ for any fixed x. Let $x - \sigma$ be introduced as a new integration variable in (51). Property (i) and an application of integration by parts to (51) show that $I(x, t)$ exists and that

$$I(x, \tau) = \int_{x-\beta}^{x-\alpha} \{F(x - \sigma) - F(x)\} d_\sigma \psi(\sigma, \tau) + o(1)$$

as $\tau \to +0$. If $|x| < l$, then, for some δ, where $0 < \delta < \eta$,

$$I(x, \tau) = -\int\limits_{|\sigma| < \delta} \{F(x) - F(x - \sigma)\} d_\sigma \psi(\sigma, \tau) + o(1)$$

as $\tau \to +0$, by virtue of property (ii) and the boundedness of $F(\sigma)$. If $F'(x)$ exists, property (iv) implies that

$$I(x, \tau) - F'(x) = \int\limits_{|\sigma| < \delta} \{F'(x) - (F(x) - F(x - \sigma))/\sigma\} \sigma d_\sigma \psi(\sigma, \tau) + o(1)$$

as $\tau \to +0$. Finally, if $\delta = \delta_\epsilon$ is chosen so small that

$$|F'(x) - (F(x) - F(x - \sigma))/\sigma| < \epsilon \text{ for } |\sigma| < \delta,$$

it is seen that

$$\limsup_{\tau\to+0} | I(x,\tau) - F'(x) | \leqq \epsilon \limsup_{\tau\to+0} \int_{|\sigma|<\delta} | \sigma | \, | d_\sigma\psi(\sigma,\tau) | .$$

Since $\epsilon > 0$ is arbitrary, property (iii) implies the assertion (52) of Lemma 4.

Proof of (XII). Let $u(x,t)$ be representable in the form (9), in terms of functions (12_1)-(12_3). It will first be shown that if $0 < s < 1$, and if $F_1'(s)$ exists, then

$$(53_1) \qquad \lim_{x\to+0} u_1(x,s) = F_1'(s).$$

The identity (16) shows that

$$-\theta_x(x,t) = \phi(x,t) + \tfrac{1}{2}\pi^{-\frac{1}{2}} o(x,t), \qquad (t > 0),$$

where $\phi(x,t)$ is the source function (10) and

$$o(x,t) = \sum_{k=1}^{\infty} t^{-3/2}\{(x+2k)\exp(-(x+2k)^2/4t) + (x-2k)\exp(-(x-2k)^2/4t)\}.$$

Since this series is uniformly convergent for $0 \leqq x \leqq x_1 < 2$ and $0 < t \leqq 1$ and since each term is 0 for $x = 0$,

$$o(x,t) \to 0, \text{ as } x \to 0, \text{ uniformly for } 0 < t \leqq 1.$$

Hence, (12_1) shows that (53_1) reduces to

$$\lim_{x\to+0} \int_0^t \phi(x, t-s)\, dF_1(s) = F_1'(s),$$

provided that this last relation is valid. To insure its validity, it is sufficient to show that the function $\psi(\sigma,\tau)$ defined by

$$\psi(\sigma,\tau) = \phi(\tau,\sigma) \qquad (0 < \tau \leqq 1, 0 < \sigma \leqq 1)$$

and $\psi(0,\tau) = 0$, possesses properties (i)-(iv), Lemma 4. (In this case, $\psi(\sigma,\tau)$ is defined, not for $| \sigma | \leqq l$ but for $0 \leqq \sigma \leqq 1$. This is irrelevant; in fact, $\psi(\sigma,\tau)$ can be defined to be 0 for negative σ.)

Property (i) is equivalent to $\lim_{x\to+0} \phi(x,t) = 0$ for $0 < t \leqq 1$, which holds, by (10). Property (ii) is equivalent to

$$\lim_{x\to+0} \int_\delta^1 | \phi_t(x,s) | \, ds = 0 \text{ for } 0 < \delta < 1.$$

Since $2\pi^{\frac{1}{2}}\phi_t(x,t) = -\tfrac{1}{2}xt^{-5/2}(3 - \tfrac{1}{2}x^2/t)\exp(-x^2/4t)$, property (ii) holds. Property (iii) is equivalent to

$$\limsup_{x\to+0}\int_0^{\eta} s\,|\,\phi_t(s,x)\,|\,ds<\infty \text{ for some } \eta>0.$$

Since $\partial(t\phi(x,t))/\partial t=\phi(x,t)+t\phi_t(x,t)$, a pair of sufficient conditions for (iii) is

$$(54)\qquad \limsup_{x\to+0}\int_0^{\infty}|\,d_s s\phi(x,s)\,|<\infty$$

and

$$(55)\qquad \limsup_{x\to+0}\int_0^{\infty}|\,\phi(x,s)\,|\,ds<\infty.$$

However, by (10), the function $2\pi^{\frac{1}{2}}\partial(t\phi(x,t))/\partial t$, being identical with $\frac{1}{2}xt^{-5/2}(\frac{1}{2}x^2-t)\exp(-x^2/4t)$, vanishes only for $t=\frac{1}{2}x^2$. Hence, by (10), the integral in (54) is $2(2\pi)^{-\frac{1}{2}}\exp((-\frac{1}{2})$ for all x; so that (54) holds. The integral in (55) is a constant multiple of

$$\int_0^{\infty} xs^{-3/2}\exp(-x^2/4s)\,ds=4\int_0^{\infty}\exp(-s^2)\,ds \qquad (\tfrac{1}{2}xs^{-\frac{1}{2}}\to s).$$

Hence, (iii) is satisfied. Finally, the integral in (iv) is

$$\int_0^{\delta} s\phi_t(x,s)\,ds=\delta\phi(x,\delta)-\int_0^{\delta}\phi(x,s)\,ds.$$

But $\delta\phi(x,\delta)\to 0$, for $\delta>0$ fixed, and

$$\int_0^{\delta}\phi(x,s)\,ds=2\pi^{-\frac{1}{2}}\int_{\gamma}^{\infty}\exp(-s^2)\,ds, \qquad (\gamma=\tfrac{1}{2}x\delta^{-\frac{1}{2}}).$$

Since $\gamma\to 0$ as $x\to 0$ for $\delta>0$ fixed, (iv) is satisfied. This completes the proof of (53_1) for any s on $0<s<1$ at which $F_1'(s)$ exists.

Next, it will be shown that

$$(53_2)\qquad \lim_{x\to+0}u_2(x,s)=0, \qquad (0<s<1).$$

This is quite on the surface, since $\theta_x(x,t)\to 0$, as $x\to 1$, holds uniformly for $0<t\leqq 1$, as can be seen from (16), by writing

$$-2\pi^{\frac{1}{2}}\theta_x(x,t)=\sum_{k=1}^{\infty}t^{-3/2}\{(x+2k)\exp(-(x+2k)^2/4t)$$
$$+(x-2k-2)\exp(-(x-2k-2)^2/4t)\}$$

(the latter series is uniformly convergent for $0<t\leqq 1$ and for x on any

closed interval not containing an even integer, while each term is 0 for $x = 1$). Hence, (53_2) follows from (12_2). Also

$$(53_3) \qquad \lim_{x \to +0} u_3(x, s) = 0 \qquad (0 < s < 1).$$

This follows from (25), in view of the uniform convergence with respect to x (for $t > 0$ fixed).

The relations (53_1), (53_2), (53_3) imply the statement of (XII) concerning (6_1). Since $\theta(x, t)$ is even, of period 2 in x (for $t > 0$ fixed), the statement concerning (6_2) follows from that concerning (6_1).

In order to prove the assertion of (XII) involving (6_3), it will be shown that

$$(56_1) \qquad \lim_{t \to +0} u_1(x, t) = 0, \qquad (0 < x < 1);$$

$$(56_2) \qquad \lim_{t \to +0} u_2(x, t) = 0, \qquad (0 < x < 1);$$

and that if $0 < x < 1$ and if the derivative $F_3'(x)$ exists, then

$$(56_3) \qquad \lim_{t \to +0} u_3(x, t) = F_3'(x).$$

The formula (12_1) can be written as

$$u_1(x, t) = -\int_0^t \theta_x(x, s)\, dF_1(t - s).$$

But (17) implies, for $0 < x < 1$,

$$u_1(x, t) = o(1) \int_0^t |\, dF_1(s) \,|, \qquad (t \to 0).$$

Consequently, (56_1) follows from (13_1). Similarly, (56_2) follows from (17) and (13_2). Thus, only (56_3) remains to be proved.

Let $0 < x < 1$ and $0 < \delta < x < 1 - \delta$. Then Lemma 2 and (25) show that, in the proof of (56_3), it can be supposed that $dF_3(s) = 0$ for $0 < s < \delta$ and $1 - \delta < s < 1$; or, more generally, that not only is (13_3) satisfied, but that $F_3(s)$ is defined and of bounded variation $0 \leqq s \leqq 1$ (and also that $F_3(0) = F_3(1)$). Then, if the identity (16) is used, (12_1) becomes

$$\tfrac{1}{2}(t\pi)^{-\frac{1}{2}} \sum_{k=-\infty}^{+\infty} \Big\{ \int_0^1 \exp(-(x - s + 2k)^2/4t)\, dF_3(s)$$

$$- \int_0^1 \exp(-(x + s + 2k)^2/4t)\, dF_3(s) \Big\}.$$

If $F_3(s)$ is defined for $-\infty < s < \infty$ by the periodicity condition $F_3(s + 1)$

$= F_3(s)$, an obvious change of variables in each of the integrals in the last series shows that

$$u_3(x,t) = (t\pi)^{-\frac{1}{2}} \int_{-\infty}^{+\infty} \exp(-(x-s)^2/4t)\, dF_3(s).$$

Hence, (56_3) will follow from Lemma 4 if it is shown that the function

$$\psi(\sigma,\tau) = \tfrac{1}{2}(\tau\pi)^{-\frac{1}{2}} \exp(-\sigma^2/4\tau)$$

has the properties (i)-(iv), with $l = \infty$.

Condition (i) is implied by $\psi(\pm\infty, \tau) = 0$ and

$$\lim_{t\to+0} \tau^{-\frac{1}{2}} \exp(-\sigma^2/4\tau) = 0, \qquad \sigma \neq 0.$$

For a fixed $\tau > 0$, the function $\psi(\sigma,\tau)$ is a monotone function of σ on each of the half-lines $-\infty < \sigma \leqq -\delta$, $\delta \leqq \sigma < \infty$ and satisfies $\psi(\pm\infty,\tau) = 0$. Hence, (ii) is a consequence of (i). As in the case treated above (in connection with (54) and (55)), a pair of sufficient conditions for (iii) is

$$\text{(57)} \qquad \limsup_{t\to+0} \int_{-\infty}^{+\infty} \tau^{-\frac{1}{2}}\, |\, d_\sigma \sigma \exp(-\sigma^2/4\tau)\, | < \infty$$

and

$$\limsup_{t\to+0} \int_{-\infty}^{+\infty} \tau^{-\frac{1}{2}} \exp(-\sigma^2/4\tau)\, d\sigma < \infty.$$

However, $\partial(\tau^{-\frac{1}{2}}\sigma \exp(-\sigma^2/4\tau))/\partial\sigma = \tau^{-\frac{1}{2}}(1 - \frac{1}{2}\sigma^2/\tau)\exp(-\sigma^2/4\tau)$ vanishes only for $\sigma^2 = 2\tau$. Hence, the integral in (57) is $2^{3/2}\exp(-\frac{1}{2})$ for all τ; so that (57) holds. The change of variables $\sigma/2\tau^{\frac{1}{2}} \to \sigma$ shows that the integral in the last formula line is the integral of $2\exp(-\sigma^2)$ over $-\infty < \sigma < \infty$. Thus, condition (iii) is satisfied. Finally, the integral in (iv) is

$$(\tau\pi)^{-\frac{1}{2}}\delta \exp(-\delta^2/4\tau) - \int_{-\delta}^{+\delta} \tfrac{1}{2}(\tau\pi)^{-\frac{1}{2}} \exp(-\sigma^2/4\tau)\, d\sigma.$$

For $\delta > 0$ fixed, the first term is $o(1)$ as $\tau \to +0$; while the change of integration variables $\sigma/2\tau^{\frac{1}{2}} \to \sigma$ shows that the integral is

$$\pi^{-\frac{1}{2}} \int_{-\infty}^{+\infty} \exp(-\sigma^2)\, d\sigma + o(1) = 1 + o(1)$$

as $\tau \to +0$. Thus, (iv) holds.

Consequently, (56_3) follows from Lemma 4. This completes the proof of (XII).

Proof of (V). The existence of the integrals (8_1)-(8_3) as Lebesgue integrals follows from the first part of the proof of (IX); cf. the remarks following the italicized assertion (X). That the sum (9) of (8_1)-(8_3) is a continuous solution of (1) on S is a consequence of (IX). Since (9) is a weighted average of the bounded functions (f_1, f_2, f_3) with non-negative weights and total weight one (cf. the maximum principle used near the end of the proof of (IV)), it follows that (9) is bounded on S. Finally, (6_1)-(6_3) holds almost everywhere by virtue of (XII). This proves (V).

Proof of (VI). It follows from the assertion (X) that, if $u(x, t)$ is a continuous solution of (1) on S and is bounded on S, then u has a representation as a sum (9) of integrals (12_1)-(12_3), where F_1, F_2, F_3 satisfy (13_1)-(13_3), respectively. On the other hand, an examination of the proof of (X) shows that the assumption that u is bounded implies that F_1, F_2, F_3 are absolutely continuous with bounded derivatives. Consequently, u has a representation as a sum (9) of integrals (8_1)-(8_3), where f_1, f_2, f_3 are bounded measurable functions. That u satisfies (6_1)-(6_3) almost everywhere follows from (V). This completes the proof of (VI).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] G. Doetsch, *Theorie und Anwendung der Laplace-Transformation,* New York, 1943.

[2] M. Gevrey, "Sur les équations aux dérivées du type parabolique," *Journal de Mathématiques,* ser. 6, vol. 9 (1913), pp. 305-471.

[3] G. Herglotz, "Ueber Potenzreihen mit positivem, reellem Teile im Einheitskreis," *Berichte über die Verhandlungen der königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig, Math.-phys. Klasse,* vol. 63 (1911), pp. 501-511.

[4] H. Lebesgue, "Sur les intégrales singulières," *Annales de la Faculté des Sciences de l'Université de Toulouse,* ser. 3, vol. 1 (1909), pp. 25-128.

[5] E. E. Levi, "Sull'equazione del calore," *Annali de Matematica,* Ser. 3, vol. 14 (1908), pp. 187-264.

[6] A. Ostrowski, "Ueber die Bedeutung der Jensenschen Formel für einige Fragen der komplexen Funktionentheorie," *Acta Litterarum ac Scientiarum* (Szeged), vol. 1 (1922-23), pp. 80-87.

[7] B. Riemann-H. Weber, *Die partiellen Differentialgleichungen der mathematischen Physik,* vol. 2 (1901), pp. 111-118.

[8] A. Wintner, "The real elliptic θ_3-function," *Duke Mathematical Journal,* vol. 15 (1948), pp. 407-411.

ON THE DISTRIBUTIONS OF THE ZEROS OF FUNCTIONS BELONGING TO CERTAIN QUASI-ANALYTIC CLASSES.*[1]

By I. I. HIRSCHMAN, JR.

1. Introduction. Let $f(t)$ be an infinitely differentiable function defined for $-\infty < t < \infty$, and let M_n be a given sequence of positive constants. We shall write $f(t) \varepsilon C\{M_n, k\}$ whenever

$$(1) \qquad | f^{(n)}(t) | \leqq A k^n M_n \qquad -\infty < t < \infty,$$

for all n and for a suitable constant A which may depend on $f(t)$.

Let $f(t) \varepsilon C\{n!, k\}$. Then $f(t)$ is the restriction to the real axis of a function which is analytic in a symmetric strip about the real axis of half width $1/k$ and is bounded in every smaller symmetric strip. Let $Z(u)$ be the number of zeros of $f(t)$ counted according to their multiplicities in the closed interval $-u \leqq t \leqq u$. It is known that if $f \not\equiv 0$ then

$$(2) \qquad \limsup_{t\to\infty} (2/\pi t) \log Z(t) \leqq k.$$

On the other hand if a distribution of zeros is given for which the limit superior in equation (2) is less than k then there exists a function $f(t) \varepsilon C\{n!, k\}$ which has precisely these zeros. See [1; vol. II, p. 111].

It is our purpose to extend such theorems to functions belonging to certain quasi-analytic classes. A typical example of our results is that if $f(t) \varepsilon C\{n!(\log [n+e])^n, k\}$ and if $Z(u)$ counts the zeros of $f(t)$ as above then

$$(3) \qquad \limsup_{t\to\infty} (2/\pi t) \log\log Z(t) \leqq k.$$

On the other hand if a distribution of zeros is given for which the limit superior in equation (3) is less than k, then there exists a function $f(t) \varepsilon C\{n!(\log [n+e])^n, k\}$ with precisely these zeros.

An essential step in our procedure is a theorem which we may illustrate by the following particular case.

Let $f(t) \varepsilon C\{n!(\log [n+e])^n, k\}$. If for an infinite sequence of intervals

$$| t - t_n | \leqq k^{-1} \exp[-\pi k' t_n/2], \qquad \lim t_n = +\infty,$$

---

* Received June 20, 1949.

[1] Research supported in part by the Office of Naval Research.

we have

$$|f(t)| \leqq \exp[-\exp\exp(\pi k' t_n/2)],$$

and if $k' > k$ then $f(t) \equiv 0$. A less precise result was demonstrated in [4] where it was shown that $f \equiv 0$ follows from the assumption

$$|f(t)| \leqq \exp[-\exp\exp(\pi k' t/2)] \qquad k' > k$$

for all large positive t. The methods of [4] are however applicable to more general quasi-analytic classes than those treated here.

2. Auxiliary functions and preliminary results. Let M_n be a set of positive constants such that $\lim n \to \infty\ M_n^{1/n} = \infty$. We set, as is customary,

$$(1) \qquad T(v) = \operatorname*{Max}_{n \geqq 0} v^n/M_n \qquad 0 \leqq v < \infty.$$

We note that

$$(2) \qquad \operatorname*{Max}_{v \geqq 0} v^n/T(v) \leqq M_n \qquad n = 0, 1, \cdots.$$

See [6]. Given $v \geqq 0$ we define $n_T(v)$ to be the largest integer n for which $v^n/M_n = T(v)$. We set

$$(3) \qquad H(v) = (2/\pi) \int_1^v \log T(t) t^{-2} dt.$$

It is well known, see [6], that the class $C\{M_n\}$ consisting of all functions $f(t)$ such that for some k, $f(t) \in C\{M_n; k\}$, is quasi-analytic if and only if $H(v) \to \infty$ as $v \to \infty$. Thus when $C\{M_n\}$ is quasi-analytic the function $\mathfrak{H}(v)$ which is inverse to $H(v)$ is well defined for $0 \leqq v < \infty$, and

$$(4) \qquad v = (2/\pi) \int_1^{\mathfrak{H}(v)} \log T(t) t^{-2} dt.$$

We shall need the following results.

Lemma 2a. *Let $M_n = n![\nu(n)]^n$ where $\nu(0) > 0$, and $\nu(n)$ is non-decreasing. If $f_1(t), f_2(t) \in C\{M_n, h\}$ then $f_1(t) \cdot f_2(t) \in C\{M_n, k\}$ for $k > h$.*

We have

$$|f_i^{(n)}(t)| \leqq A_i h^n M_n \qquad -\infty < t < \infty;\ i = 1, 2;\ n = 0, 1, \cdots.$$

Using Leibnitz's rule we find that

$$[f_1(t) f_2(t)]^{(n)} = \sum_{j=0}^{n} C_j^n f_1^{(j)}(t) f_2^{(n-j)}(t).$$

$$|[f_1(t) f_2(t)]^{(n)}| \leqq A_1 A_2 \sum_{j=0}^{n} C_j^n h^j j! [\nu(j)]^j h^{n-j} (n-j)! [\nu(n-j)]^{n-j}$$

$$\leqq A_1 A_2 \sum_{j=0}^{n} n! [\nu(n)]^j [\nu(n)]^{n-j} h^n, \leqq A_1 A_2 (n+1) h^n n! [\nu(n)]^n.$$

Our lemma immediately follows.

By $f(x) \sim g(x)$ and $f(x) \leqq \sim g(x)$ we mean respectively that

$$\lim_{x\to\infty} f(x)/g(x) = 1 \text{ and that } \limsup_{x\to\infty} f(x)/g(x) \leqq 1.$$

LEMMA 2b. *Let* $M_n = n![\nu(n)]^n$ *where* $\nu(x)$ *is continuously differentiable for* $0 \leqq x < \infty$. *If* $\nu(0) = 1$, $\nu'(x) \geqq 0$, $x\nu'(x)/\nu(x) = o(1)$ *as* $x \to \infty$ *then*

A. $H(ax) \sim H(x)$ $\qquad a > 0, x \to \infty.$

B. $H(x_1x_2) \leqq \sim H(x_1) + H(x_2)$ $\qquad x_1, x_2 \to \infty.$

C. $\theta H(x) \leqq \sim H(x^\theta)$ $\qquad 0 < \theta < 1, x \to \infty.$

D. $\log \mathfrak{H}(\lambda x) \geqq \sim \lambda \log \mathfrak{H}(x)$ $\qquad \lambda > 1, x \to \infty.$

By definition

$$\log T(u) = \operatorname*{Max}_{n=0,1,\ldots} [n \log u - \log \Gamma(n+1) - n \log \nu(n)].$$

Treating n as a continuous variable and noting that $\log \Gamma(n+1) \sim (n+\frac{1}{2}) \log n - n$; $(d/dn) \log \Gamma(n) = \log n + o(1)$, we have

$$\begin{aligned}(d/dn)&[n \log u - \log \Gamma(n+1) - n \log \nu(n)] \\ &= \log u - \log n - \log \nu(n) - n\nu'(n)/\nu(n) + o(1).\end{aligned}$$

This implies that $n_T(u)\nu[n_T(u)] \sim u$, $\log T(u) \sim n_T(u)$, and, combining these, that $\log T(u) \sim u/\nu[n_T(u)]$. Noting that ν and n_T are both non-decreasing functions, it is apparent that there exists a continuous non-decreasing function $\psi(u)$, for which we may take $\psi(0) > 0$, such that $\operatorname{Log} T(u) \sim u/\psi(u)$.

We may now demonstrate conclusion **B**. The other results may be dealt with in a similar manner. We have

$$\begin{aligned} H(x_1x_2) &\sim (2/\pi) \int_1^{x_1x_2} [u\psi(u)]^{-1} du \qquad x_1, x_2 \to \infty \\ &\sim (2/\pi) \int_1^{x_1} [u\psi(u)]^{-1} du + (2/\pi) \int_{x_1}^{x_1x_2} [u\psi(u)]^{-1} du. \end{aligned}$$

Making the change of variable $x_1 v = u$ in the second integral we have

$$H(x_1x_2) \sim (2/\pi) \int_1^{x_2} [u\psi(u)]^{-1} + (2/\pi) \int_1^{x_2} [v\psi(x_1v)]^{-1} dv,$$

or, since $\psi(u)$ is non-decreasing,

$$\begin{aligned} H(x_1x_2) &\leqq \sim (2/\pi) \int_1^{x_1} [u\psi(u)]^{-1} du + (2/\pi) \int_1^{x_2} [v\psi(v)]^{-1} dv \\ &\leqq \sim H(x_1) + H(x_2). \end{aligned}$$

as desired.

We leave to the reader to verify that under the slightly stronger assumption $x\nu'(x)/\nu(x) = 0\,(\log x)^{-1}$ we have $\log T(u) \sim u/\nu(u)$.

Let us agree to write $\exp_m x$ and $\log_m x$ for

$$\exp\exp\cdots\exp x \quad\text{and}\quad \log\log\cdots\log x,$$

respectively. For m a positive integer we define

$$N_n{}^{(m)} = n!, \qquad 0 \leqq n < \exp_m 1; \qquad N_n{}^{(m)} = n![\log n \log_2 n \cdots \log_m n]^n, \quad n \geqq \exp_m 1.$$

The class $C\{N_n{}^{(m)}\}$ is called the m-th logarithmic class. It is easily seen from the above remark that for $C\{N_n{}^{(m)}\}$ we have

$$\log T(v) \sim v/\log v \log_2 v \cdots \log_m v; \qquad H(v) \sim (2/\pi)\log_{m+1} v;$$
$$\log_{m+1} \mathfrak{H}(v) \sim (\pi/2)v.$$

See also [4].

Definition 2c. A quasi-analytic class $C\{M_n\}$ will be said to be regular if $M_n = n!\nu(n)^n$ where $\nu(x)$ satisfies the assumptions of Lemma 2b.

We shall make fundamental use of the following theorem due to Gorny [3] and Cartan [2].

THEOREM 2d. *If*

1. $|F^{(n)}(t)| \leqq P_n \qquad |t - t'| \leqq 1/\lambda;\ n = 0, 1, \cdots,$
2. $P_\tau \geqq \tau! P_0 \lambda^\tau,$

then

$$|F^{(k)}(t')| \leqq 2e^k P_0{}^{1-k\tau^{-1}} P_\tau{}^{k\tau^{-1}} \qquad 0 \leqq k \leqq \tau.$$

3. Functions which are small on a sequence of intervals.

THEOREM 3a. *Let $C\{M_n\}$ be a regular class and let H and $\mathfrak{H}$ be defined as in Section* 2. *If $f(t) \varepsilon C\{M_n; k\}$, if for any sequence of pairs $\{t_n, T_n\}$, $0 < t_1 < t_2 < \cdots$, $\lim t_n = \infty$, $0 \leqq T_n \leqq t_n$, we have in each of the intervals $|t - T_n| \leqq 1/k\nu[\mathfrak{H}(k't_n)]$ the inequality $|f(t)| \leqq O(1)\exp[-\mathfrak{H}(k't_n)]$, and if $k' > k$, then $f(t) \equiv 0$.*

We note that it is no restriction to assume that $f(t) \geqq 0$, for by Lemma 2a the function $f_1(t) = f(t)^2$ also belongs to $C\{M_n; k_1\}$ for any $k_1 > k$, and $f_1 \equiv 0$ entails $f \equiv 0$.

Next it is no restriction to suppose that $A = 1$, $k = 1$, $k' > 1$, $O(1) = 1$, since we could consider instead of $f(t)$ the function $Bf(tb)$ for suitable constants B and b.

Let us choose constants η_n, $0 \leqq \eta_n < 1$, such that $\tau_n = \mathfrak{H}[k't_n] + \eta_n$ is an integer. We apply the theorem of Gorny and Cartan to $f(t)$ with

$$\tau = \tau_n, \qquad t' = T_n; \qquad \lambda = \nu[\mathfrak{H}(k't_n)] \leqq \nu(\tau_n);$$
$$P_0 = \exp[-\mathfrak{H}(k't_n)] \leqq \exp[-\tau_n + 1]; \qquad P_\tau = \tau_n![\nu(\tau_n)]^{\tau_n}.$$

It is immediately verifiable that $P_\tau \geqq \tau!P_0\lambda^\tau$, so that the assumption of Theorem 2d is satisfied. We have

$$|f^{(k)}(T_n)| \leqq 6e^{2k}e^{-\tau_n}\{\tau_n![\nu(\tau_n)]^{\tau_n}\}^{k/\tau_n} \leqq 6e^{2k}e^{-\tau_n}[\tau_n\nu(\tau_n)]^k, \qquad 0 \leqq k \leqq \tau_n.$$

Let us define $\phi_n(w) = \int_0^\infty e^{-wt}f(T_n - t)\,dt$, $w = u + iv$. For $\epsilon > 0$ we have $|\phi_n(\epsilon \pm iv)| \leqq M_0/\epsilon$, $0 \leqq v < \infty$. We assert that given θ, $0 < \theta < 1$, the inequality $|\phi_n(\epsilon \pm iv)| \leqq 2/\epsilon T(v)$, $1 \leqq v \leqq \tau_n{}^\theta$, will hold for all n sufficiently large. We integrate the formula defining $\phi_n(w)$ by parts r times to obtain

$$\phi_n(w) = [\sum_{k=0}^{r-1} (-1)^k f^{(k)}(T_n)w^{-k-1}] + (-1)^r w^{-r}\int_0^\infty e^{-wt}f^{(r)}(T_n - t)\,dt.$$

This implies that

$$|\phi_n(\epsilon \pm iv)| \leqq [\sum_{k=0}^{r-1} |f^{(k)}(T_n)|\,v^{-k-1}] + M_r v^{-r}/\epsilon.$$

It is clearly sufficient to prove that

$$\sum_{k=0}^{n_T(v)-1} |f^{(k)}(T_n)|\,v^{-k-1} \leqq 1/\epsilon T(v) \qquad 1 \leqq v \leqq \tau_n{}^\theta,$$

where $n_T(v)$ is defined as in Section 2. Now $1/T(v) \geqq e^{-v}$, $n_T(v) \leqq v + 1$, so that remembering that $v \geqq 1$, $T(v)\ \varepsilon \uparrow$, it is enough to show that

$$\sum_{0 \leqq k \leqq \tau_n{}^\theta} 6e^{2k}e^{-\tau_n}[\tau_n\nu(\tau_n)]^k \leqq e^{-\tau_n{}^\theta}/\epsilon$$

for n sufficiently large. Since $\nu(x) = O(x^\delta)$ for any $\delta > 0$ this is obvious.

By the principle of harmonic majoration we now have, for n sufficiently large,

$$\log|\phi_n(1 + \epsilon)| \leqq (2/\pi)\{\int_0^1 \log(M_0/\epsilon)[1 + t^2]^{-1}dt$$
$$+ \int_1^{\tau_n{}^\theta} [\log(2/\epsilon) - \log T(t)][1 + t^2]^{-1}dt$$
$$+ \int_{\tau_n{}^\theta}^\infty \log(M_0/\epsilon)[1 + t^2]^{-1}dt\}, \leqq -H[\tau_n{}^\theta] + C,$$

where C is a constant independent of n. Using Lemma 2b, and remembering that $\theta < 1$ is arbitrary we see that

$$(1) \qquad \log |\phi_n(1+\epsilon)| \leqq \sim - k't_n \qquad n \to \infty.$$

Now $\phi_n(1+\epsilon) = \int_0^\infty e^{-(1+\epsilon)t} f(T_n - t)\,dt = e^{-(1+\epsilon)T_n} \int_{-T_n}^\infty e^{-u} f(-u)\,du$. Since $f \geqq 0$, $\phi_n(1+\epsilon) \geqq e^{-(1+\epsilon)t_n} \int_0^\infty e^{-u} f(-u)\,du$. Thus

$$(2) \qquad \log \phi_n(1+\epsilon) \geqq \sim -(1+\epsilon)t_n \qquad n \to \infty,$$

if $f \not\equiv 0$. Equations (1) and (2) are in contradiction if ϵ is sufficiently small and this implies that $f(t) \equiv 0$, which is what we wished to prove.

THEOREM 3b. *Let* ϵ, $0 < \epsilon \leqq 1$, *be given; then Theorem* 3a *remains valid if the inequalities hold in the intervals* $|t - T_n| \leqq \epsilon\{k\nu[\mathfrak{H}(k't_n)]\}^{-1}$ *instead of in the intervals* $|t - T_n| \leqq \{k\nu[\mathfrak{H}(k't_n)]\}^{-1}$.

We shall merely sketch the modifications necessary to prove this slightly more general result. Let $\delta_n = 2\epsilon/k\nu[\mathfrak{H}(k't_n)]$, and let N be an integer so large that $(2N+1)\epsilon \geqq 1$. We consider instead of the function $f(u)$ the sequence of functions

$$f^*_n(u) = [f(u)f(u+\delta_n)f(u-\delta_n)f(u+2\delta_n)\cdots f(u+N\delta_n)f(u-N\delta_n)]^2.$$

If k_1 is any constant $k < k_1 < k'$, then by Lemma 2a, we can choose A_1 independently of n so that

$$|[f^*_n(u)]^{(m)}| \leqq A_1 k_1^m M_m, \qquad -\infty < u < \infty;\ m = 0, 1, \cdots.$$

Further $|f^*_n(u)| \leqq O(1)\exp[-\mathfrak{H}(k't_n)]$ for $|u - T_n| \leqq 1/k\nu[\mathfrak{H}(k't_n)]$. As before we may suppose $k_1 = 1$, etc. We define

$$\phi_n(w) = \int_0^\infty e^{-wu} f^*_n(T_n - u)\,du.$$

Our assumptions lead, by the arguments used previously, to the relation

$$\lim_{n\to\infty} \int_0^\infty f^*_n(-u)\,du = 0,$$

which is impossible unless $f(u) \equiv 0$.

If we apply Theorem 3b, with $\epsilon = 1$ and $T_n = t_n$, to the first logarithmic class, then we obtain the theorem quoted in the introduction. If we apply Theorem 3b to the analytic class and then make an exponential change of variable, we obtain the following result.

COROLLARY 3c. *Let* $f(z)$ *be analytic and bounded in the half plane* $\mathrm{Rl}\, z \geqq 0$ *and let* $\epsilon > 0$ *be given. If for an infinite sequence of values* $0 < x_1 < x_2 < \cdots, \lim_{n\to\infty} x_n = \infty$, *we have in each of the intervals*

$$(1-\epsilon)x_n \leqq x \leqq (1+\epsilon)x_n$$

the inequalities $|f(x)| \leqq e^{-x_n{}^k}$ *and if* $k > 1$, *then* $f(z) \equiv 0$.

This special case can serve as an indication of the degree of precision of our results; see [5; Chapter VII].

4. The zeros of functions belonging to regular classes.

LEMMA 4a. *If*

1. $|F^{(n)}(t)| \leqq Ak^n M_n \qquad -\infty < t < \infty;\ n = 0, 1, \cdots.$

2. $F(t)$ *has* N *zeros, counted according to their multiplicities, in the interval* $a \leqq t \leqq b$, $b - a = 1/\lambda$. *Then for every* n, $0 \leqq n \leqq N$, *we have* $|F(t)| \leqq A(k/\lambda)^n M_n/n!$ $a \leqq t \leqq b$.

Although this is doubtless well known, we shall, for the sake of completeness, give the proof. Let $\zeta_1, \zeta_2, \cdots, \zeta_n$ be zeros of $F(t)$ with possible repetitions in the case of multiplicities. Consider a value t, $a \leqq t \leqq b$. If $F(t) = 0$ there is nothing to prove. If $F(t) \neq 0$ we form the function

$$F(u) \prod_{i=1}^{n} (t - \zeta_i) - F(t) \prod_{i=1}^{n} (u - \zeta_i).$$

This function is zero for $u = t, \zeta_1, \zeta_2, \cdots, \zeta_n$, and therefore, by Rolle's theorem, its n-th derivative must vanish at some point $u = \xi$, $a < \xi < b$; i. e.

$$F(t) = F^{(n)}(\xi)[n!]^{-1} \prod_{i=1}^{n} (t - \xi_i).$$

Since $|t - \zeta_i| \leqq 1/\lambda$, $|F^{(n)}(\xi)| \leqq Ak^n M_n$, we obtain the desired inequality.

THEOREM 4b. *Let* $C\{M_n\}$ *be a regular quasi-analytic class and let* $f(t) \in C\{M_n, k\}$, $f(t) \neq 0$. *If* $Z(t)$ *is the number of zeros of* $f(u)$ *in the interval* $-t \leqq u \leqq t$ *counted according to their multiplicities, then*

$$\limsup_{t\to\infty} t^{-1} H[Z(t)] \leqq k. \tag{1}$$

Because of conclusion **A** of Lemma 2b it will be enough to show that

$$\limsup_{t\to\infty} t^{-1} H[Z_+(t)] \leqq k \tag{2}$$

where $Z_+(t)$ is the number of zeros of $f(u)$ in the interval $0 \leqq u \leqq t$. If equation (2) is not true, there exists a constant $k' > k$, and an infinite sequence of values $0 < t_1 < t_2 \cdots$, $\lim t_n = \infty$, such that

$$(3) \qquad H(Z_+(t_n)) \geqq k't_n \qquad n = 1, 2, \cdots,$$

or, equivalently, such that

$$(4) \qquad Z_+(t_n) \geqq \mathfrak{H}(k't_n) \qquad n = 1, 2, \cdots.$$

By conclusion **D** of Lemma 2b we see that if we choose k'', $k < k'' < k'$, then for all sufficiently large n

$$Z_+(t_n) \geqq 5kt_n\mathfrak{H}(k''t_n)\nu[\mathfrak{H}(k''t_n)].$$

We divide the interval $0 \leqq t \leqq t_n$ into subintervals of length $1/\{4k\nu[\mathfrak{H}(k''t_n)]\}$ beginning at t_n and proceeding to the left. There may be a fractional interval abutting zero. There are less than $5kt_n\nu[\mathfrak{H}(k''t_n)]$ of these intervals and therefore one of them, which when n is large cannot be the one abutting zero, must contain more than $\mathfrak{H}(k''t_n)$ zeros. Let τ_n be the smallest integer greater than or equal to $\mathfrak{H}(k''t_n)$ and let T_n be the center of the interval in question. Applying Lemma 4a we find that if

$$(5) \qquad |t - T_n| \leqq \{8k\nu[\mathfrak{H}(k''t_n)]\}^{-1},$$

we have $|f(t)| \leqq A\{k\nu(\tau_n)/4k\nu[\mathfrak{H}(k''t_n)]\}^{\tau_n}$. But $\lim_{n\to\infty} \nu(\tau_n)/\nu[\mathfrak{H}(k''t_n)] = 1$, which implies that for n sufficiently large $|f(t)| \leqq \exp(-\tau_n)$, or

$$(6) \qquad |f(t)| \leqq \exp[-\mathfrak{H}(k''t_n)].$$

Equations (5) and (6) and Theorem 3b show that $f(t) \equiv 0$ contrary to our assumption. Thus equation (3) must hold and our theorem is proved.

COROLLARY 4c. *Let* $f(t) \in C\{N_n^{(m)}, k\}$ *and let* $Z(t)$ *be the number of zeros of* $f(u)$ *in the closed interval from* $-u$ *to* u, *counted according to their multiplicities. Then we have* $\limsup_{t\to\infty} (2/\pi t)\log_{m+1} Z(t) \leqq k$.

5. On the construction of functions with prescribed zeros. We know from Section 2 that if $C\{M_n\}$ is a regular quasi-analytical class, then $\log T(u) \sim u/\psi(u)$, where $\psi(u)$ is continuous, non-decreasing and $\psi(0) > 0$. We set

$$(1) \qquad t = (2/\pi) \int_1^u [u\psi(u)]^{-1} du,$$

$$(2) \qquad R(u) = \log \mathfrak{H}^*(u), \qquad \rho(u) = \psi[\mathfrak{H}^*(u)].$$

We have

$$(3) \qquad \psi(\exp R(u)) = \rho(u), \qquad R(u) = (\pi/2)\int_1^u \rho(x)\,dx.$$

The first of these relations is immediate; the second may be demonstrated by differentiating equation (1) and the first equation of (2) to obtain $R'(u) = (\pi/2)\rho(u)$, and then noting that $R(0) = 0$.

Let $\epsilon > 0$ be given. It will be convenient to write λ for $1 + \epsilon$.

Let S_z be the strip in the z plane, $z = x + iy$ defined by the inequality $|y| < \pi/2$ for $-\infty < x < \infty$, and let S_w be the curvilinear strip in the w plane, $w = u + iv$ defined by the inequalities $|v| \leqq 1/\lambda\rho(0)$ for $-\infty < u \leqq 0$, $|v| \leqq 1/\lambda\rho(u)$ for $0 \leqq u < \infty$. Let $\chi(w)$ be the function which maps S_w onto S_z conformally and is normalized by the conditions $\chi(0) = 0$, $\chi'(0) > 0$.

Let $F(z)$ be analytic and bounded in S_z.

We define

$$(4) \qquad f(w) = [4\rho(0)^2 + w^2]^{-1}\exp\{-d(\lambda)\exp[\chi(w)/\lambda]\}F[\chi(w)],$$

where $d(\lambda) = \exp(4\pi/\lambda)/\cos(\pi/2\lambda)$.

THEOREM 5a. *Let* $C\{M_n\}$ *be a regular quasi-analytic class and let* $f(w)$ *be defined by equation* (4), *then* $f \varepsilon C\{M_n, \lambda^4\}$.

We shall throughout the present discussion denote by A any constant depending only on $C\{M_n\}$, and the bound of $|F(z)|$ in S_z. Let us define

$$(5) \qquad \phi(t) = \int_{-\infty}^{\infty} e^{itu} f(u)\,du.$$

We begin by establishing that

$$(6) \qquad |\phi(t)| \leqq A\exp[-|t|/\lambda\psi(|t|)], \qquad -\infty < t < \infty.$$

It is sufficient to consider the case $t \geqq 0$. We deform, as we evidently may, the path of integration in the integral (5) until it becomes the upper boundary P of S_w. If P' and P'' are the parts of P in the left and right half planes respectively, then

$$\phi(t) = \int_{P'} e^{itw} f(w)\,dw + \int_{P''} e^{itw} f(w)\,dw = I' + I''.$$

Now $|f(w)| \leqq A(1 + u^2)^{-1}$ on P' so that $|I'| \leqq A\exp(-t/\lambda\rho(0))$ and, a fortiori, $|I'| \leqq A\exp[-t/\lambda\psi(t)]$, $t \geqq 0$. In order to estimate I'', we appeal to Ahlfors' distortion theorem, see [1], to show that for w on P'' we have $\mathrm{Rl}\,\chi(w) \geqq \lambda R(u) - 4\pi$. This implies that for P on W'' $\exp\{-d(\lambda)\exp[\chi(w)/\lambda]\} \leqq \exp[-\exp R(u)]$. If $\sigma(v)$ is the arc length along P'' between the abscissas $u = 0$ and $u = v$, then $\sigma(u) \leqq u + (\lambda\rho(0))^{-1}$, and

$$|I''| \leqq A \int_0^\infty \exp[-t/\lambda\rho(u)]\exp[-\exp R(u)](1+u^2)^{-1}d\sigma(u).$$

Let us put $t = \lambda\rho(\tau)\exp R(\tau)$. We split the above integral into two integrals corresponding to the ranges of integration $(0,\tau)$ and (τ,∞). Since $\rho(u)$ and $R(u)$ are non-decreasing we find by simple estimations that $|I''| \leqq A\exp[-\exp R(\tau)]$. Thus to prove that $|I''| \leqq A\exp[-t/\lambda\psi(t)]$ we need only verify that for τ sufficiently large we have $\exp[-\exp R(\tau)] \leqq \exp[-t/\lambda\psi(t)]$, which, inserting for t its expression in terms of τ, will be true if $\psi(\exp R(\tau)) \geqq \rho(\tau)$. The validity of this inequality follows from equation (3). Thus equation (6) has been established.

Since $\log T(u) \sim u/\psi(u)$ we have

$$|\phi(t)| \leqq A/[T(|t|)]^{1/\lambda^2}, \qquad -\infty < t < \infty.$$

By the Fourier inversion formula

$$f(u) = (2\pi)^{-1}\int_{-\infty}^{\infty} e^{-itu}\phi(t)\,dt; \qquad f^{(n)}(u) = (2\pi)^{-1}\int_{-\infty}^{\infty} e^{-itu}(-it)^n\phi(t)\,dt.$$

Hence

$$|f^{(n)}(u)| \leqq A\int_0^\infty |t|^n[T(t)]^{-1/\lambda^2}dt.$$

We split this integral into two integrals corresponding to the ranges $(0,1)$ and $(1,\infty)$. The first of these integrals is $O(1)$ as $n\to\infty$. Now

$$\int_1^\infty |t|^n[T(t)]^{-1/\lambda^2}dt = \int_1^\infty |t|^{-2}\{|t|^{(n+2)\lambda^2}/T(t)\}^{1/\lambda^2}dt.$$

Let n_λ be the smallest integer greater than or equal to $(n+2)\lambda^2$. Then by (2) of Section 2 we have

$$\operatorname*{Max}_{1\leqq t<\infty} t^{(n+2)\lambda^2}/T(t) \leqq \operatorname*{Max}_{0\leqq t<\infty} t^{n_\lambda}/T(t) \leqq M_{n_\lambda}.$$

Hence

$$\int_1^\infty |t|^n[T(t)]^{-1/\lambda^2}dt \leqq A(n_\lambda!)^{1/\lambda^2}[\nu(n_\lambda)]^{n_\lambda/\lambda^2}.$$

By Stirling's formula we have $(n_\lambda!)^{1/\lambda^2} \leqq A\lambda^{3n}n!$. It may be easily deduced from the relation $x\nu'(x)/\nu(x) = o(1)$ that $\nu(ax) \sim \nu(x)$ for any $a > 0$ so that

$$[\nu(n_\lambda)]^{n_\lambda/\lambda^2} \leqq A\lambda^n[\nu(n)]^n.$$

Combining our inequalities we have

$$|f^{(n)}(u)| \leqq An![\nu(n)]^n\lambda^{4n} \qquad -\infty < u < \infty;\ n = 0, 1, \cdots,$$

as desired.

THEOREM 5b. *Let $C\{M_n\}$ be a regular quasi-analytic class, let an arbitrary set of real zeros with associated multiplicities be given, and let $Z(u)$ be the number of these zeros counted according to their multiplicities in the closed interval from $-u$ to u. If $\limsup_{t\to\infty} t^{-1}H[Z(t)] < k$, then there exists a function $f \varepsilon C\{M_n, k\}$ which has precisely these zeros.*

Because of Lemma 2a we may restrict ourselves to the case where the zeros are non-negative. It will also be sufficient to suppose $k < 1$ and to prove that $f \varepsilon C\{M_n, k'\}$ for any pre-assigned $k' > 1$.

Choose $\epsilon > 0$ so small that if $\lambda = (1 + \epsilon)$ then $\lambda^4 \leqq k'$, and let χ, R, etc. be defined as in Theorem 5a. Let $0 \leqq \zeta_1 \leqq \zeta_2 \leqq \zeta_3 \leqq \cdots$, be the given zeros arranged in order of increasing value and repeated according to their multiplicities. Let $x_k = \chi(\zeta_k)$ $n = 1, 2, \cdots$, and let $X(u)$ be the number of x_k's in the closed interval from 0 to u. By Ahlfors' distortion theorem we have $X[\lambda R(u) - 4\pi] \leqq Z(u)$. By assumption we have for all u sufficiently large $Z(u) \leqq \mathfrak{H}(u)$. Thus if $r = R(u)$ we have for r sufficiently large $X[\lambda r - 4\pi] \leqq e^r$, i. e., $X(r) = O(e^{-r/\lambda})$, $r \to \infty$. It follows, see [1], that there exists a function $F(z)$, analytic and bounded in S_z which has the zeros x_k with the correct multiplicities and such that these are the only zeros of $F(z)$.

If now $f(u)$ is defined as in equation (4), then f will have just the prescribed zeros and, by Theorem 5a, $f \varepsilon C\{M_n, k'\}$.

HARVARD UNIVERSITY AND WASHINGTON UNIVERSITY.

BIBLIOGRAPHY.

1. L. Bieberbach, *Lehrbuch der Funktionen-theorie*, Leipzig (1931).

2. H. Cartan, *Sur les classes de fonctions définies par des inégalités portant sur les dérivées successives.* Actualités Scientifiques et Industrielles 867, Paris (1940).

3. A. Gorny, "Contributions à l'étude des fonctions dérivables d'une variable réelle," *Acta Mathematica*, vol. 71 (1939), pp. 317-358.

4. I. I. Hirschman, Jr., "On the behavior of Fourier transforms at infinity and on quasi-analytic classes of functions," *American Journal of Mathematics*, vol. 72 (1950), pp. 200-213.

5. N. Levinson, *Gap and density theorems*, New York (1939).

6. S. Mandelbrojt, "Analytic functions and classes of infinitely differentiable functions," *The Rice Institute Pamphlet*, vol. 29 (1942), pp. 1-142.

BESSEL FUNCTION APPROXIMATIONS.*

By R. S. Phillips and Henry Malin.

In this paper[1] we have obtained the following bounds for the logarithmic derivatives of the modified Bessel and Hankel functions:

$$\begin{aligned}\phi_n\{v, [n(n+1)]^{\frac{1}{2}}\} &< v^{-1}I_n'(v)/I_n(v) < \phi_n(v, n),\\ \phi_n(v, n) &< -v^{-1}K_n'(v)/K_n(v) < \phi_n\{v, [n(n-1)]^{\frac{1}{2}}\}\end{aligned} \tag{1}$$

for $v > 0$ and $n \geqq 1$, where

$$\phi_n(v, \alpha) = (n/v^2)[1 + (v/\alpha)^2]^{\frac{1}{2}}. \tag{2}$$

These inequalities result from a study of the behavior of the functions

$$\begin{aligned}X_n(v) &= v^{-1}I_n'(v)/I_n(v) - \phi_n(v, \alpha);\\ Y_n(v) &= -v^{-1}K_n'(v)/K_n(v) - \phi_n(v, \alpha)\end{aligned} \tag{3}$$

by means of the Ricatti differential equation which they satisfy.

Theorem 1. *If $0 < \alpha \leqq n$, then X_n is negative and monotonic increasing, and its graph approaches the v-axis from below; if $n < \alpha < [n(n+1)^2(n+2)]^{1/4}$, then X_n has exactly one maximum and no minima, and its graph thereafter approaches the v-axis from above; if $\alpha \geqq [n(n+1)^2(n+2)]^{1/4}$, then X_n is positive and monotonic decreasing, and its graph approaches the v-axis from above.*

This theorem is valid for all integers $n \geqq 0$ and $v > 0$.

Making use of the fact that $I_n(v)$ satisfies Bessel's differential equation, it is readily found that X_n satisfies the differential equation

$$\begin{aligned}dZ/dv = &-vZ^2 - 2/v\{n[1 + (v/\alpha)^2]^{\frac{1}{2}} + 1\}Z + v^{-1}\\ &-(n/\alpha)^2 v^{-1} - (n/\alpha^2)v^{-1}[1 + (v/\alpha)^2]^{-\frac{1}{2}}.\end{aligned} \tag{4}$$

One can likewise write down the series and asymptotic expansions for $X_n(v)$ from known expansions for $I_n(v)$; they are:

---

* Received June 27, 1949.

[1] The work of this paper was done under funds granted to New York University by the Watson Laboratories of the Air Material Command, Red Bank, N. J., under contract W28-099-ac-170.

$$(5) \qquad X_n(v) = \{[2(n+1)]^{-1} - n/(2\alpha^2)\} - \{[8(n+1)^2(n+2)]^{-1} - n/(8\alpha^4)\}v^2 + \cdots,$$

$$(6) \qquad X_n(v) = [1 - (n/\alpha)]v^{-1} - (1/2)v^{-2} + \cdots.$$

The differential equation (4) associates a direction to every point in the (v, Z)-plane. The point and associated direction are known as a line-element. In particular we will be interested in the locus of all points where the slope of the line element is zero. Setting dZ/dv equal to zero results in a quadratic equation in Z. The roots of this quadratic in Z are two functions of v. Let $U(v)$ be the greater and $L(v)$ be the smaller of these two roots. The curves representing these two functions divide the half-plane $v \geqq 0$ into three regions: the region above the U-curve in which dZ/dv is negative; the region between the U- and L-curves in which dZ/dv is positive; and the region below the L-curve in which dZ/dv is again negative.

Most of the proof is concerned with an investigation of the U- and L-curves. The behavior of the X_n curve for small v is known by the series (5). We know therefore to which of the three regions X_n belongs initially. Recall that X_n belongs to the family of Z-curves. It follows that X_n will be monotonic increasing (decreasing) until it intersects with zero slope, a U- or L-curve, after which it is monotonic decreasing (increasing); etc. The known behavior of the X_n at infinity enables us to complete the theorem. It turns out that the X_n behave like $U = U(v; \alpha, n)$.

We begin, then, by setting dZ/dv equal to zero in equation (4) and find that the explicit expressions for U and L are

$$(7) \quad U, L = v^{-1}[-v^{-1}(ns+1) \pm \{1 + (n^2+1)v^{-2} + (n/s)(\alpha^{-2} + 2v^{-2})\}^{\frac{1}{2}}],$$

where $s = [1 + (v/\alpha)^2]^{\frac{1}{2}}$; the plus sign goes with U, and the minus sign with L. It is clear from equation (7) that U and L are always real valued so that their graphs actually do divide the right half-plane into the three regions described above.

One sees from equation (7) that $L(v)$ is the sum of negative monotonic increasing terms and hence is itself negative monotonic increasing. Furthermore, $L(v)$ has a pole at the origin. It follows from (5) that the graph of X_n starts out above that of $L(v)$. Furthermore if it should ever touch the $L(v)$-curve it would cross (having a zero slope at this point) and enter a region of negative slope. It could never again intersect the $L(v)$-curve since $L(v)$ is a monotonic increasing function. The value of X_n at this point of intersection would thereafter be an upper bound for X_n. Since this value is necessarily less than zero, this would be contrary to the fact that X_n approaches zero asymptotically as $v \to \infty$. Consequently the graph of X_n lies above that of $L(v)$ for all $v \geqq 0$.

The quadratic equation in Z obtained from equation (4) by setting dZ/dv equal to zero can be rewritten as a cubic function of s.[2] The result is

$$(8) \qquad F(s) \equiv \alpha^2 Z^2 s^3 + 2nZs^2 - [\alpha^2 Z^2 + 1 - 2Z - n^2/\alpha^2]s + n/\alpha^2 = 0.$$

To a given value of Z and a root $s \geqq 1$, there is a $v \geqq 0$ such that either $U(v) = Z$ or $L(v) = Z$. (This value of Z need not of course be assumed by the function X_n.) Since $F(-\infty) = -\infty$, $F(0) = n/\alpha^2$, and $F(\infty) = \infty$, it follows for a given Z, $F(s)$ has at most two roots for which $s \geqq 1$. Since $L(v)$ takes on all negative values, it follows that for $Z < 0$, one of these roots and only one necessarily corresponds to a point on $L(v)$. Hence in the region where $U(v)$ is negative, it must be monotonic increasing. On the other hand, for positive Z, either zero, one, or two roots can lie on $U(v)$. Hence in the region where $U(v)$ is positive it is either monotonic or has a single maximum (since, as is readily shown, the graph of $U(v)$ approaches the v-axis asymptotically as $v \to \infty$).

To conclude our description of $U(v)$ we make use of the series and asymptotic expansions of $U(v)$ which are readily obtainable from equation (7).

$$\begin{aligned}(9) \quad U(v) = \{[2(n+1)]^{-1} - n/(2\alpha^2)\} \\ - \{[8(n+1)^3]^{-1} - (n(n+2))/[8(n+1)\alpha^4]\}v^2 \\ + \{[16(n+1)^5]^{-1} - n/[16(n+1)^3\alpha^4] \\ - [n(n+3)]/[16(n+1)\alpha^6]\}v^4 + \cdots,\end{aligned}$$

$$(10) \quad U(v) = [1 - (n/\alpha)]v^{-1} - [1 - n/(2\alpha)]v^{-2} + \cdots.$$

For $\alpha \leqq n$, $U(v)$ is negative for v near zero and also for sufficiently large v; thus $U(v)$ is always negative and monotonic increasing. For $n < \alpha < [n(n+1)^2(n+2)]^{1/4}$, $U(v)$ has a positive slope for small v and is positive as $v \to \infty$, which implies that it has a single maximum and no minima. Finally for $\alpha \geqq [n(n+1)^2(n+2)]^{1/4}$, $U(v)$ has a negative slope for small values of v and is positive as $v \to \infty$, and therefore is always positive and monotonic decreasing.

---

[2] The change of variable $s^2 = 1 + (v/\alpha)^2$ establishes a 1-1 relationship between the two half-planes $s \geqq 1$ and $v \geqq 0$. The strip $0 \leqq s < 1$ does not, of course, correspond to any part of the real v-plane. Curves in the s-half-plane go over into curves in the v-half-plane under a horizontal stretching process; the property of being monotonic and the property of having a horizontal tangent are preserved by the transformation. It is for this reason that we shall translate freely these properties proved in the s-variable into the v-variable without further comment.

We now return to the function X_n. A comparison of the series expansion for X_n and U [equations (5) and (9)] shows that for $\alpha \leqq [n(n+1)^2(n+2)]^{1/4}$ the graph of X_n starts below that of $U(v)$. Now the graph of X_n cannot intersect that of $U(v)$ from below at a point at which $U(v)$ is increasing, because at such a point the slope of X_n would be zero (by definition of U). For $\alpha \leqq n$, U is always monotonic increasing so that X_n remains between L and U, that is in the positive slope region. Consequently in this case X_n is negative and monotonic increasing.

For $n < \alpha < [n(n+1)^2(n+2)]^{1/4}$, X_n can intersect U only at a point at which U has a non-positive slope. At such a point it must necessarily cross to the region above U, since U is either decreasing or at its maximum, whereas X_n has a zero slope and can only decrease by crossing over to the region above U. Further, it is clear that X_n *must* cross into this region since it approaches the v-axis from above and hence eventually is decreasing. Once in the region above U, X_n must remain there for all larger v, since U is thereafter a monotonic decreasing function and X_n would necessarily have a zero slope at any point of intersection. Thus X_n is increasing at the start, has its maximum at the point of intersection with U and is thereafter decreasing, approaching the v-axis asymptotically from above.

Finally, for $\alpha > [n(n+1)^2(n+2)]^{1/4}$, X_n is greater than U at the start and since U is monotonic decreasing for $v > 0$, it follows, as above, that X_n must remain above U for all $v > 0$. Hence in this case X_n is positive and monotonic decreasing. This is likewise true when the inequality is replaced by an equality since the functions X_n are continuous in α. This concludes the proof of Theorem 1.

Upper and lower bounds of the functions $v^{-1}I_n'(v)/I_n(v)$ and $[v^{-1}I_n'(v)/I_n(v)]'$ can now be readily deduced. If $\alpha^2 = n(n+1)$, the function $X_n = 0$ for $v = 0$ and is thereafter positive; for $\alpha = n$, X_n is always negative. If $\alpha = [n(n+1)^2(n+2)]^{1/4}$, then X_n' is negative for $v > 0$; whereas if $\alpha = n$, X_n' is positive for $v > 0$. These results are summarized in terms of $\phi_n(v, \alpha)$ [see (2)] by means of the following corollary:

COROLLARY 1. 1. *For all $v > 0$ and $n \geqq 1$,*

$$\phi_n\{v, [n(n+1)]^{1/2}\} < v^{-1}I_n'(v)/I_n(v) < \phi_n(v, n),$$
$$\phi_n'(v, n) < [v^{-1}I_n'(v)/I_n(v)]' < \phi_n'\{v, [n(n+1)^2(n+2)]^{1/4}\}.$$

We next proceed to prove a theorem similar to Theorem 1 for the function $Y_n(v)$ related as in (3) to the modified Hankel function $K_n(v)$.

THEOREM 2. *If $\alpha \leqq [n(n-1)^2(n-2)]^{1/4}$, then Y_n is negative and* *monotonic increasing, and its graph approaches the v-axis from below; if* $[n(n-1)^2(n-2)]^{1/4} < \alpha < n$, *then* Y_n *has a single minimum and no maxima and its graph thereafter approaches the v-axis from below; if* $\alpha \geqq n$, *then* Y_n *is positive and monotonic decreasing, and its graph approaches the v-axis from above.*

This theorem is valid for all $v > 0$ and $n \geqq 0$.

The series and asymptotic expansions for Y_n are:

$$
\begin{aligned}
&Y_0 = -[v^2 \log(\gamma v/2)]^{-1} + \cdots, \\
(11)\qquad &Y_1 = -\log(\gamma v/2) - (2\alpha^2)^{-1} \cdots, \\
&Y_n = \{[2(n-1)]^{-1} - n/(2\alpha^2)\} \\
&\qquad\qquad - \{[8(n-1)^2(n-2)]^{-1} - n/\alpha^4\}v^2 + \cdots
\end{aligned}
$$

for $n > 1$, where $C = \log \gamma = 0.577 \cdots$ is Euler's constant.

$$(12)\qquad Y_n = [1 - n/\alpha]v^{-1} + (1/2)v^{-2} + \cdots$$

for $n \geqq 0$ and large v. Further $Y_n(v)$ satisfies the following differential equation:

$$
\begin{aligned}
(13)\qquad dZ/dv = vZ^2 &+ (2/v)\{n[1 + (v/\alpha)^2]^{\frac{1}{2}} - 1\}Z - v^{-1} \\
&+ (n/\alpha)^2 v^{-1} - (n/\alpha^2)v^{-1}[1 + (v/\alpha)^2]^{-\frac{1}{2}}.
\end{aligned}
$$

We shall first show that $Y_0(v)$ is a positive monotonic decreasing function of v for all $v > 0$. It follows from (11) that Y_0 is positive for small v; it is further analytic in v for $v > 0$. If Y_0 were ever to become negative, there would exist a $v = v_1$ such that $Y_0(v_1) = 0$. For such a $v = v_1$, equation (13) shows that dY_0/dv becomes $-v_1^{-1} < 0$. Hence, although Y_0 could presumably become zero and take on negative values, it could never thereafter become positive. However, by the asymptotic expansion (12), Y_0 approaches the v-axis from above as $v \to \infty$. This behavior is consistent only with the assumption that $Y_0 > 0$ for all $v > 0$.

To go on to the monotonicity of Y_0, we see from equation (11) that dY_0/dv is negative for v small. Differentiating equation (13) we get ($n = 0$)

$$(14)\qquad d^2Y_0/dv^2 = [2vY_0 - (2/v)]dY_0/dv + Y_0^2 + (2/v^2)Y_0 + v^{-2}.$$

For any $v = v_1$ at which $dY_0/dv = 0$, equations (13) and (14) show that d^2Y_0/dv^2 reduces to $2[Y_0(v_1)]^2 > 0$. Hence if dY_0/dv were ever to vanish, it must change from negative to positive values. That is to say if dY_0/dv ever becomes positive, it must remain so. This is inconsistent with the asymptotic behavior of Y_0 as $v \to \infty$. Hence dY_0/dv is negative for all $v > 0$.

We next consider the general case $n \geqq 1$. As in Theorem 1, we are interested in the locus of points where $dZ/dv = 0$. This locus is obtained as the solution of a quadratic in Z and defines the two functions $U(v)$ and $L(v)$ which are respectively the greater and the smaller of the roots of this quadratic. From equation (13), the explicit expressions for U and L are:

$$(15) \quad U, L = v^{-2}(1 - ns) \pm \{1 + (1 + n^2)v^{-2} - (n/s)(\alpha^{-2} + 2v^{-2})\}^{\frac{1}{2}},$$

where $s = [1 + (v/\alpha)^2]^{\frac{1}{2}}$; the plus sign goes with U and the minus sign with L.

It is no longer true that U and L are defined for all v and n. One can therefore not expect U and L to have the simple properties that they had in Theorem 1; it is precisely this fact that complicates the proof of the present theorem. If U and L are everywhere defined, the half-plane $v > 0$ breaks up into three regions in each of which the sign of dZ/dv is everywhere the same: the region above the U-curve in which dZ/dv is positive, the region between the U- and L-curves in which dZ/dv is negative, and the region below the L-curve in which dZ/dv is positive. When U and L are not everywhere defined, the U- and L-curves have two branches; in the regions interior to the U-, L-curves, dZ/dv is negative; in the region exterior to the U-, L-curves, dZ/dv is positive.

One can readily obtain the series and asymptotic expansions of U from equation (15). They are

$$(16) \quad \begin{aligned} U_1 &= v^{-1} - (2\alpha^2)^{-1} + \cdots, \\ U_n &= \{[2(n-1)]^{-1} - n/(2\alpha^2)\} \\ &\quad - [8(n-1)]^{-1}\{(n-1)^{-2} - n(n-2)/\alpha^4\}v^2 + \cdots \ (n > 1) \end{aligned}$$

for small v, and

$$(17) \quad U_n = [1 - (n/\alpha)]v^{-1} + [1 - n/(2\alpha)]v^{-2} + \cdots, \qquad (n \geqq 1),$$

for large v.

We see from equation (15) that for those values of v for which L exists, L is the sum of two negative terms and is therefore itself negative. Furthermore, L has a pole at the origin. It follows that any solution curve which is bounded from below at the origin (hence Y_n in particular) starts out above the L-curve.

As in Theorem 1, we shall again study the U- and L-curves by means of the cubic function of s obtained from equation (13) by setting $dZ/dv = 0$, namely,

$$(18) \quad F(s) \equiv \alpha^2 Z^2 s^3 + 2nZs^2 - \{\alpha^2 Z^2 + 2Z + 1 - (n^2/\alpha^2)\}s - n/\alpha^2 = 0.$$

To each value of Z and a root $s \geqq 1$ of $F(s) = 0$, there is a v such that either $U(v) = Z$ or $L(v) = Z$.

We shall first consider the case $\alpha \geqq n$. It is easy to see that the radicand in equation (15) is positive since $1 > (sn)^{-1} \geqq n/(s\alpha^2)$ and $1 + n^2 \geqq 2n > 2n/s$ for $s > 1$. It follows that U and L are defined (and single valued) for all $s > 1$. Furthermore, for $Z > 0$ and $\alpha \geqq n$ we see that the coefficients of $F(s)$ have the signs $(+ + - -)$ in the order of decreasing powers of s. Thus there is just one variation in sign, and by Descartes' rule of signs there is precisely one root for $s > 0$. Now the U-curve has a positive ordinate and negative slope at $s = 1$ and approaches the v-axis from above as $v \to \infty$. If U were to have a minimum, there would be values of Z such that $U(v) = Z$ for at least two different values of v, and hence two different positive values of s. Thus $F(s) = 0$ for more than one positive value of s. Since this is impossible, U can have no minimum and we conclude that U is positive and monotonic decreasing for $\alpha \geqq n$.

From the expansion of Y_n [equation (11)] we see that the graph of Y_n starts off positive with a negative slope and hence lies between the U-curve and the L-curve for small $v > 0$. Since U is monotonic decreasing, the graph of Y_n could cross that of U with zero slope. If it did so, however, it would thereafter be in the region of positive slope above the graph of U and hence would be bounded away from zero. This is contrary to the fact that Y_n approaches zero from above as v approaches infinity [see equation (12)]. Finally if Y_n were ever zero, the asymptotic behavior of Y_n indicates that it would be zero an even number of times. This means that, at the points of crossing, its slope would alternate in sign. This is impossible since for $Y_n = Z = 0$, equation (13) reduces to

$$dZ/dv = -v^{-1}[1 - (n/\alpha)^2] - (n/\alpha^2)v^{-1}[1 + (v/\alpha)^2]^{-\frac{1}{2}} < 0 \quad \text{for all} \quad v > 0.$$

It follows that Y_n is positive and monotonic decreasing for $\alpha \geqq n$.

We have as a corollary to the first part of the theorem that

$$-v^{-1}K'_n(v)/K_n(v) > (n/v^2)[1 + (v/n)^2]^{\frac{1}{2}} \tag{19}$$

for all $v > 0$. Hence if we define the auxiliary function

$$\psi(v) = (n/v^2)\{[1 + (v/n)^2]^{\frac{1}{2}} - [1 + (v/\alpha)^2]^{\frac{1}{2}}\} \tag{20}$$

it follows from (19) that, for all $v > 0$ and all $\alpha > 0$,

$$Y_n(v) > \psi(v). \tag{21}$$

We now consider the case $\alpha < n$. In this case U and L need not be defined for all $v \geqq 0$. It will be convenient to consider U and L as functions[3] of s; that is, as roots of $F(s) = 0$.

$$U, L = [\alpha^2(s^2 - 1)]^{-1}\{1 - ns \pm [1 + n^2 + \alpha^2(s^2 - 1) - (n/s)(1 + s^2)]^{\frac{1}{2}}\}.$$

For s small and positive, the radicand is negative so that the curves do not exist in this region. Each curve has its only positive singularity at $s = 1$. This can be seen from the expansion of U about $s = 1$

$$U = (1 - n)/[\alpha^2(s - 1)] + \cdots \qquad (s < 1, n > 1)$$

$$U = \alpha^{-1}[2(s - 1)]^{-\frac{1}{2}} + \cdots \qquad (s > 1, n = 1).$$

Now for $Z > 0$ the coefficients of $F(s)$ have the signs $(+ + \pm -)$ and hence $F(s)$ has one and only one positive real root. If $Z = 0$, $F(s) = 0$ for $s = [n(1 - \sigma)]^{-1}$, where we have set $\sigma = (\alpha/n)^2$. For any $Z \geqq 0$ the value of s satisfying $F(s) = 0$ makes $U(s) = Z$. Hence, the graph of $Z = U(s)$ crosses the s-axis at P_1: $([n(1 - \sigma)]^{-1}, 0)$ and crosses every line $Z =$ constant > 0 just once. If we try to trace the curve $Z = U(s)$ we find first that for $n > 1$ it approaches the line $s = 1$ asymptotically from the left; it cannot cross the Z-axis, so that it must double back on itself; since it crosses each horizontal line $(Z > 0)$ just once it can have no minimum; $F(s)$ is quadratic in Z and hence the curve cannot double back on itself more than once for $s < 1$; for $s > 1$ it cannot double back on itself above the s-axis, since this would imply a root $L > 0$, which is impossible for $s > 1$ (i. e. $v > 0$). It follows that for $Z > 0$ and $n > 1$ the curve must look like the sketches in Fig. a and b.

If $n = 1$, the graph of $Z = U$ approaches the line $s = 1$ asymptotically from the right, and since it crosses the s-axis at $(1 - \sigma)^{-1} > 1$, it must decrease monotonically until it reaches this point. (See Fig. c; $Z > 0$).

We shall study U for $Z < 0$ by means of the auxiliary function $\psi(v)$ defined by equation (20). At $v = 0$, $U = [2(n - 1)]^{-1} - n/(2\alpha^2)$, $\psi = (2n)^{-1} - n/(2\alpha^2)$, and $L \to -\infty$ as $v \to 0+$; therefore ψ lies between U and L at $v = 0$. In order to determine the points at which the graph of ψ intersects that of U or L we go back to the equation defining U and L; that is, we insert $Z = \psi(v)$ in equation (13), setting $dZ/dv = 0$. One obtains by direct substitution of (20) into (13)

$$(2n/v^3)\{[1 + (v/\alpha)^2]^{\frac{1}{2}} - [1 + (v/n)^2]^{\frac{1}{2}}\} - [n/(v\alpha^2)][1 + (v/\alpha)^2]^{-\frac{1}{2}} = 0.$$

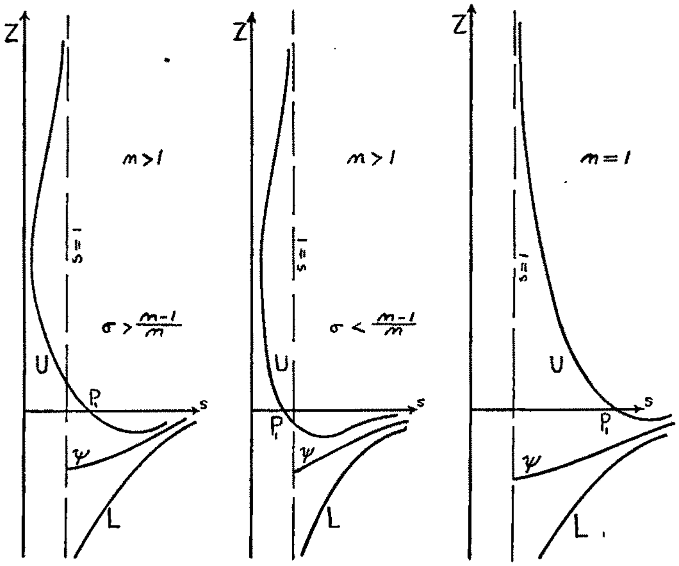
Solving for the real roots in v (that is $s \geqq 1$), we get $(v/\alpha)^2 = 4\sigma/(1 - 4\sigma)$, $(\sigma = (\alpha/n)^2 > 0)$, if and only if $\sigma < 1/4$. For this value of v,

---

[3] It should be noted that for $0 < s < 1$, L can be positive; whereas for $1 \leqq s$ (i. e. $v > 0$) L must be negative where it exists.

$$\psi = -[n/(2\alpha^2)][1-4\sigma]^{\frac{1}{2}} \text{ and } s = [1-4\sigma]^{-\frac{1}{2}}.$$

It follows that for $n/2 \leqq \alpha < n$, the graph of ψ intersects neither the U- nor the L-curves for $v \geqq 0$, since in this case $\sigma \geqq \frac{1}{4}$. This implies that U and L are not connected and hence must exist and be single-valued for all $v > 0$. The graph of L starts at minus infinity and approaches the v-axis asymptotically from below. Hence L must assume all negative values of Z at least once. For a given value of $Z < 0$, U can therefore correspond to at most two roots of the cubic $F(s) = 0$. Since the U-curve crosses into the region $Z < 0$ from above and approaches the s-axis asymptotically from below, U can have but a single minimum for α in the range $n/2 \leqq \alpha < n$. It is clear that $U > \psi > L$. The situation is therefore represented by the sketches in Fig. a, b, c.[4]



(a) (b) (c)

For $\alpha < n/2$, the function ψ has just one point in common with U or L $(v \geqq 0)$, namely P_2:

$$([1-4\sigma]^{-\frac{1}{2}}, -(2\sigma n)^{-1}[1-4\sigma]^{\frac{1}{2}})$$

---

[4] It should be noted that for $s \geqq 1$, U is single-valued because of the existence of L below it. For $s < 1$, however, the extension of U to the left of $s = 1$ (the lower part of which would properly be called L) is double-valued because of the existence of the upper branch of U which approaches $s = 1$ asymptotically from the left (if $n > 1$).

(in the (sZ)-plane). Now

$$\psi - (U + L)/2 = \{n[1 + (v/n)^2]^{\frac{1}{2}} - 1\}v^{-2} > 0;$$

thus ψ is always greater than the average of U and L $(v \geqq 0)$, and hence the graph of ψ certainly lies above that of L. Therefore it intersects the graph of U once $(v \geqq 0)$ and thereafter lies above both the U- and the L-curves.

We now wish to show that to the left of P_2, that is, for $0 < s \leqq [1 - 4\sigma]^{-\frac{1}{2}}$, U has at most one minimum and no maximum. We show first that in this range, $F(s)$ has precisely one root for $-(2n\sigma)^{-1}[1 - 4\sigma]^{\frac{1}{2}} < Z < 0$. It follows that there is one value of s for each Z in this range for which $U(s) = Z$. We define $G(Z)$ as follows:

$$\begin{aligned} G(Z) \equiv {} & [(4\sigma^2 n^2)(1 - 4\sigma)^{-3/2}]Z^2 + 2[n - (1 - 4\sigma)^{\frac{1}{2}}][1 - 4\sigma]^{-1}Z \\ & + [n(1 - \sigma) - (1 - 4\sigma)^{\frac{1}{2}}][n\sigma(1 - 4\sigma)^{\frac{1}{2}}]^{-1} = F[(1 - 4\sigma)^{-\frac{1}{2}}]. \end{aligned}$$

$G(Z)$ is thus a parabolic function of Z, and

$$G(0) = [n(1 - \sigma) - (1 - 4\sigma)^{\frac{1}{2}}][n\sigma(1 - 4\sigma)^{\frac{1}{2}}]^{-1} > 0,$$

since $[1 - 4\sigma]^{\frac{1}{2}}/n \leqq [1 - 4\sigma]^{\frac{1}{2}} < 1 - 2\sigma < 1 - \sigma$. Further since P_2 lies on U, $G(-[1 - 4\sigma]^{\frac{1}{2}}/(2n\sigma)) = 0$. The minimum of the parabolic function, $G(Z)$, is

$$Z_m = - [1 - 4\sigma]^{\frac{1}{2}}/(2n\sigma) \cdot [n - (1 - 4\sigma^{\frac{1}{2}}]/(2n\sigma) < - [1 - 4\sigma]^{\frac{1}{2}}/(2n\sigma),$$

since as above $(n - [1 - 4\sigma]^{\frac{1}{2}}) > 2n\sigma$. Hence Z_m lies below P_2. Now $G(Z) = 0$ for $Z = - [1 - 4\sigma]^{\frac{1}{2}}/(2n\sigma)$ and the minimum of $G(Z)$ occurs for $Z < - [1 - 4\sigma]^{\frac{1}{2}}/(2n\sigma)$; it follows that $G(Z)$ is greater than 0 for all Z in the range $- [1 - 4\sigma]^{\frac{1}{2}}/(2n\sigma) < Z \leqq 0$. Returning to the function $F(s)$, we have

$$F(0) = - n^2/\alpha^2 < 0 \qquad (Z \text{ arbitrary});$$

$$F([1 - 4\sigma]^{-\frac{1}{2}}) \equiv G(Z) > 0 \qquad (\text{for } - [1 - 4\sigma]^{\frac{1}{2}}/(2n\sigma) < Z \leqq 0).$$

Hence the equation $F(s) = 0$ has an odd number of roots for $0 \leqq s \leqq [1 - 4\sigma]^{-\frac{1}{2}}$ and fixed Z, $(- [1 - 4\sigma]^{\frac{1}{2}}/(2n\sigma) < Z \leqq 0)$. If we can show that the graph of $F(s)$ has no point of inflection in this region it will follow that there is exactly one root. Now $F''(s) = 0$ for $s = - 2/(3n\sigma Z)$. For $Z > - [1 - 4\sigma]^{\frac{1}{2}}/(2n\sigma)$, i. e. $-Z < [1 - 4\sigma]^{\frac{1}{2}}/(2n\sigma)$, this value of $s > (4/3)(1 - 4\sigma)^{-\frac{1}{2}} > (1 - 4\sigma)^{-\frac{1}{2}}$; thus the point of inflection is to the right of P_2.

As we have seen, the U-curve enters the lower half of the (s, Z)-plane by crossing the s-axis at the point P_1. Since $F(s)$ has precisely one root to the left of P_2 for each negative Z above P_2, the U-curve must proceed downward without a maximum or a minimum in the region R above and to the left of P_2. It must pass out of R at P_2 or to the left of P_2, since P_2 would otherwise lie on L. If it passes out of R to the left of P_2 it can not again enter R, because if it were to do so $F(s)$ would have more than one root for some value of Z in R. On the other hand, the curve eventually passes through P_2, so that it must have a minimum below P_2. Furthermore it can have no other extremum below P_2, since $F(s)$ can have at most three roots for any value of Z. Finally for v in the range $0 \leqq v \leqq 2\alpha[\sigma/(1-4\sigma)]^{\frac{1}{2}}$ (i. e. $1 \leqq s \leqq [1-4\sigma]^{-\frac{1}{2}}$), the curve L exists below ψ; if follows that $U(v)$ is single-valued in this range. Thus $U(v)$ is single-valued and has at most one minimum but no maxima to the left of P_2. There are thus only three possible descriptions of $U(v)$ to the left of P_2. $U(v)$ is monotonic decreasing; $U(v)$ has a single minimum and no maximum; $U(v)$ is monotonic increasing. The latter case occurs only if $U(v)$ is initially increasing, i. e. if $\alpha \leqq [n(n-1)^2(n-2)]^{1/4}$.

The proof of the theorem for $\alpha < n$ is now straightforward. When $[n(n-1)^2(n-2)]^{1/4} < \alpha < n$ the graph of Y_n starts off below the U-curve with a negative slope. In the upper half-plane, U decreases monotonically until its graph intersects the v-axis. The graph of Y cannot intersect that of U above the v-axis, for if it did so it would enter a region of positive slope and remain bounded away from the v-axis; this is contrary to the behavior of Y_n at infinity. Furthermore, the graph of Y_n cannot intersect (with zero slope) that of U where U is increasing. On the other hand, since Y_n is eventually increasing, it must enter a region of positive slope at some time. Now $Y_n > \psi > L$. Hence the graph of Y_n intersects the U-curve below the v-axis at a point at which U is decreasing. Thereafter Y_n cannot enter a region of negative slope. For if it intersected the graph of U again, the intersection would have to occur at a point of increasing slope for U (to the left of P_2 if $\alpha < n/2$). In this case Y_n would become trapped between ψ and a monotonically increasing portion of U; it could never again enter a region of increasing slope. This is contrary to the behavior of Y_n at infinity.

For $\alpha \leqq [n(n-1)^2(n-2)]^{1/4}$, the graphs of both Y_n and U start out with positive slopes and with Y_n above U. In this case U is monotonic increasing at least up to the point P_2. Hence if Y_n ever entered a region of decreasing slope, it would have to intersect the U-curve to the left of P_2.

As before Y_n would thereafter be trapped between ψ and a monotonically increasing portion of U. In this case Y_n would remain in a region of decreasing slope and therefore be bounded away from the v-axis; this is contrary to the behavior of Y_n at infinity. It follows that Y_n is monotonic increasing for all $v > 0$. This concludes the proof of Theorem 2.

Theorem 2 furnishes us with some useful bounds on the function $-v^{-1}K'_n(v)/K_n(v)$ and its derivative. For $\alpha^2 = n(n-1)$, Y_n is initially zero and is negative thereafter; for $\alpha = n$, Y_n is always positive. For $\alpha = [n(n-1)^2(n-2)]^{\frac{1}{4}}$, dY_n/dv is positive $(v > 0)$; for $\alpha = n$, dY_n/dv is negative $(v > 0)$. We have thus proved

COROLLARY 2.1. *For all* $n \geqq 1$ *and all* $v > 0$,

$$\phi_n(v, n) < -v^{-1}K_n'(v)/K_n(v) < \phi_n\{v, [n(n-1)]^{\frac{1}{2}}\},$$

$$\phi_n'\{v, [n(n-1)^2(n-2)]^{\frac{1}{4}}\} < [-v^{-1}K_n'(v)/K_n(v)]' < \phi_n'(v, n).$$

Combining the results of the previous corollaries to Theorems 1 and 2,

$$-v^{-1}K_n'(v)/K_n(v) > \phi_n(v, n) > v^{-1}I_n'(v)/I_n(v).$$

In other words, $I_n'(v)/I_n(v) + K_n'(v)/K_n(v) < 0$, and hence:

COROLLARY 2.2. *For all* $v \geqq 0$ *and* $n \geqq 0$,

$$[I_n(v)K_n(v)]' < 0.$$

The case $n = 0$ requires a special and somewhat tedious argument. For the details of the proof we refer the reader to the original report of which this paper is a part.[5]

THE UNIVERSITY OF SOUTHERN CALIFORNIA,
COLLEGE OF THE CITY OF NEW YORK,
NEW YORK UNIVERSITY.

---

[5] See pp. 54-57 of 'A Helical Wave Guide' by Phillips and Malin, New York University Mathematical Research group, report No. 170-3, Army Air Forces, Watson Laboratories.

ON THE APPROXIMATION OF IRRATIONAL NUMBERS BY THE CONVERGENTS OF THEIR CONTINUED FRACTIONS, II.*

By Alfred Brauer and Nathaniel Macon.

Introduction. This paper is a continuation of our previous paper with the same title [this Journal, vol. 71 (1949), pp. 349-361]. The enumeration of the chapters, theorems, and equations will be continued here. We are now able to improve the results of Chapter 4, in which we obtained lower bounds for the sums of two, three, four, and eight consecutive λ_n and used the results to obtain a lower bound for $\underline{\lim}\,(\sum_{i=0}^{m-1}\lambda_i)/m$. In this paper, these results are not only improved, but in addition we obtain the best possible bound for the sum of k consecutive λ_n for any given k, and the best possible bound for $\underline{\lim}\,(\sum_{i=0}^{m-1}\lambda_i)/m$.

If

$$f_{-2}=1,\ f_{-1}=0,\ f_0=1,\ f_1=1,\ f_2=2,\ f_3=3,\ f_4=5,\cdots \tag{45}$$

are the numbers of Fibonacci, then we set

$$F_k=\sum_{i=0}^{k-1} f_{i-1}/f_i. \tag{46}$$

It will be shown that for every irrational number ξ, for every k, and for every n, we have $\sum_{\kappa=0}^{k-1}\lambda_{n+\kappa}>k+2F_k$, but for every integer k and for every $\epsilon>0$ it is possible to construct irrational numbers ξ for which $\sum_{\kappa=0}^{k-1}\lambda_{n+\kappa}<k+2F_k+\epsilon$, for infinitely many values of n.

A theorem of Hurwitz [10] can be formulated as follows:

$$\lim_{m\to\infty}\,(\sum_{i=0}^{m-1}\lambda_i)/m=5^{\frac{1}{2}}$$

for $\xi=(1+5^{\frac{1}{2}})/2$ and equivalent numbers ξ.

The following generalization of this theorem will be proved. For every irrational number ξ, we have $\underline{\lim}\,(\sum_{i=0}^{m-1}\lambda_i)/m \geqq 5^{\frac{1}{2}}$.

---

* Received May 2, 1949. Presented to the American Mathematical Society, December 29, 1949.

5. Some lemmas. Let $\xi = [q_0, q_1, \cdots, q_n, \cdots]$ be a given irrational number. We shall consider the sum $\lambda_n + \lambda_{n+1} + \cdots + \lambda_{n+k-1}$ for any given n and any $k > 1$.

Let t be an integer such that $1 \leqq t \leqq k$, and let $q_{n+t} + 1 = q^*_{n+t}$. Consider the irrational number $\xi^* = [q_0, q_1, \cdots, q_{n+t-1}, q^*_{n+t}, q_{n+t+1}, \cdots]$, and denote by $\lambda^*_{n+\kappa}$, $\xi^*_{n+\kappa}$, $\phi^*_{n+\kappa}$ the values corresponding to $\lambda_{n+\kappa}$, $\xi_{n+\kappa}$, $\phi_{n+\kappa}$, respectively, for ξ^*.

LEMMA 2. *We have*

$$\sum_{\kappa=1}^{k} (\xi^*_{n+\kappa} - \xi_{n+\kappa}) > \tfrac{1}{2}. \tag{47}$$

Proof. It follows from (6) that

$$\xi_{n+\kappa} = [q_{n+\kappa}, q_{n+\kappa+1}, \cdots]. \tag{48}$$

Hence $\xi^*_{n+\kappa} = \xi_{n+\kappa}$, $(\kappa = t+1, t+2, \cdots)$, and it is sufficient to show that

$$\sum_{\kappa=t}^{1} (\xi^*_{n+\kappa} - \xi_{n+\kappa}) > \tfrac{1}{2}. \tag{49}$$

Now

$$\xi^*_{n+t} - \xi_{n+t} = 1, \tag{50}$$

and

$$\xi^*_{n+t-1} - \xi_{n+t-1} = q_{n+t-1} + 1/(\xi_{n+t} + 1) - q_{n+t-1} - 1/\xi_{n+t}$$
$$= -1/\xi_{n+t}(\xi_{n+t} + 1).$$

Since $\xi_{n+t} > 1$, we have

$$-\tfrac{1}{2} < \xi^*_{n+t-1} - \xi_{n+t-1} < 0. \tag{51}$$

Moreover, $\xi^*_{n+\kappa-1} - \xi_{n+\kappa-1} = 1/\xi^*_{n+\kappa} - 1/\xi_{n+\kappa} = (\xi_{n+\kappa} - \xi^*_{n+\kappa}) / (\xi_{n+\kappa}\xi^*_{n+\kappa})$, $(\kappa = t-1, t-2, \cdots, 1)$. Hence the left member of (49) is an alternating series with decreasing terms. Thus it follows from (50) and (51) that $\sum_{\kappa=t}^{1} (\xi^*_{n+\kappa} - \xi_{n+\kappa}) > 1 - \tfrac{1}{2} = \tfrac{1}{2}$.

LEMMA 3. *We have*

$$\sum_{\kappa=1}^{k} (\phi^*_{n+\kappa} - \phi_{n+\kappa}) > -\tfrac{1}{2}. \tag{52}$$

Proof. It follows from (7) that $\phi_{n+\kappa} = [0, q_{n+\kappa-1}, q_{n+\kappa-2}, \cdots, q_1]$. Hence, $\phi^*_{n+\kappa} = \phi_{n+\kappa}$, $(\kappa = 1, 2, \cdots, t)$, and we only need to show that

$$\sum_{\kappa=t+1}^{k} (\phi^*_{n+\kappa} - \phi_{n+\kappa}) > -\tfrac{1}{2}. \tag{53}$$

Now $\phi^*_{n+t+1} - \phi_{n+t+1} = 1/(q_{n+t} + 1 + \phi_{n+t}) - 1/(q_{n+t} + \phi_{n+t}) = -1/(q_{n+t} + 1 + \phi_{n+t})(q_{n+t} + \phi_{n+t})$. Hence

$$(54) \qquad \phi^*_{n+t+1} - \phi_{n+t+1} > -\tfrac{1}{2}.$$

In addition, for $\kappa = t+1, t+2, \cdots$,

$$\phi^*_{n+\kappa} - \phi_{n+\kappa} = (\phi^*_{n+\kappa} + q_{n+\kappa}) - (\phi_{n+\kappa} + q_{n+\kappa}) = 1/\phi^*_{n+\kappa+1} - 1/\phi_{n+\kappa+1}$$
$$= (\phi_{n+\kappa+1} - \phi^*_{n+\kappa+1})/\phi_{n+\kappa+1}\phi^*_{n+\kappa+1}.$$

Since $0 < \phi_{n+\kappa+1}\phi^*_{n+\kappa+1} < 1$, the left member of (53) is an alternating series with decreasing terms. Thus (53) follows from (54).

LEMMA 4. *For every $k > 1$, we have* $\sum_{\kappa=0}^{k-1} (\lambda^*_{n+\kappa} - \lambda_{n+\kappa}) > 0$.

Proof. By Lemmas 2 and 3, we have $\sum_{\kappa=0}^{k-1} (\lambda^*_{n+\kappa} - \lambda_{n+\kappa}) = \sum_{\kappa=1}^{k} (\xi^*_{n+\kappa} - \xi_{n+\kappa}) + \sum_{\kappa=1}^{k} (\phi^*_{n+\kappa} - \phi_{n+\kappa}) > \tfrac{1}{2} - \tfrac{1}{2} = 0$. By successive application of Lemma 4, we obtain

LEMMA 5. *Let* $\xi = [q_0, q_1, \cdots, q_n, q_{n+2}, \cdots, q_{n+k}, q_{n+k+1}, \cdots]$ *be an arbitrary irrational number. If we set*

$$(55) \qquad \bar{\xi} = [q_0, q_1, \cdots, q_n, 1, 1, \cdots, 1, q_{n+k+1}, q_{n+k+2}, \cdots],$$

then $\sum_{\kappa=0}^{k-1} \lambda_{n+\kappa} \geqq \sum_{\kappa=0}^{k-1} \bar{\lambda}_{n+\kappa}$ *where the* $\bar{\lambda}_{n+\kappa}$ *are the* λ*-values associated with* $\bar{\xi}$.

LEMMA 6. *Let x be the finite continued fraction* $[q_0, q_1, \cdots, q_n, 1, 1, \cdots, 1]$ *where the last k partial quotients are each unity. If the complete quotients are denoted by x_v, then* $\sum_{\kappa=1}^{k} x_{n+\kappa} = k + F_k$, *where F_k is defined by* (45) *and* (46).

Proof. It follows from the recursion formula for the convergents that

$$(56) \qquad x_{n+k-i} = f_{i+1}/f_i, \qquad (i = 0, 1, \cdots, k-1).$$

Hence $\sum_{\kappa=1}^{k} x_{n+\kappa} = \sum_{i=0}^{k-1} f_{i+1}/f_i = k + \sum_{i=0}^{k-1} f_{i-1}/f_i = k + F_k$.

LEMMA 7. *Let $\bar{\xi}$ again denote the continued fraction* (55). *Then the complete quotients satisfy the inequality* $\sum_{\kappa=1}^{k} \bar{\xi}_{n+\kappa} > k + F_k$.

Proof. From (48) we obtain here $\bar{\xi}_{n+k-i} = [1, 1, \cdots, 1, \bar{\xi}_{n+k+1}]$, $(i = 0,$

$1, \cdots, k-1)$, where the first $i+1$ partial quotients are 1's. It follows from the recursion formula for the continued fraction and from (56) that

$$(57) \quad \bar{\xi}_{n+k-i} = (f_{i+1}\bar{\xi}_{n+k+1} + f_i)/(f_i\bar{\xi}_{n+k+1} + f_{i-1}), \qquad (i = 0, 1, \cdots, k-1).$$

Differentiating with respect to $\bar{\xi}_{n+k+1}$, we have $d\bar{\xi}_{n+k-i}/d\bar{\xi}_{n+k+1} = (f_i f_{i+1}\bar{\xi}_{n+k+1} + f_{i-1}f_{i+1} - f_i f_{i+1}\bar{\xi}_{n+k+1} - f_i^2)/(f_i\bar{\xi}_{n+k+1} + f_{i-1})^2$. Using the formula $f_{i-1}f_{i+1} - f_i^2 = (-1)^{i+1}$ for the Fibonacci numbers, we obtain

$$d\bar{\xi}_{n+k-i}/d\bar{\xi}_{n+k+1} = (-1)^{i+1}/(f_i\bar{\xi}_{n+k+1} + f_{i-1})^2.$$

Hence $\sum_{i=0}^{k-1} (d\bar{\xi}_{n+k-i}/d\bar{\xi}_{n+k+1}) = \sum_{i=0}^{k-1} (-1)^{i+1}/(f_i\bar{\xi}_{n+k+1} + f_{i-1})^2$. The right member is a finite alternating series with decreasing terms. Since the first term is negative, the sum is negative. Hence $\sum_{\kappa=1}^{k} \bar{\xi}_{n+\kappa}$ is a decreasing function of $\bar{\xi}_{n+k+1}$. Therefore, by (57),

$$\begin{aligned} \sum_{\kappa=1}^{k} \bar{\xi}_{n+\kappa} &> \lim_{\xi_{n+k+1}\to\infty} \sum_{i=0}^{k-1} (f_{i+1}\bar{\xi}_{n+k+1} + f_i)/(f_i\bar{\xi}_{n+k+1} + f_{i-1}) \\ &= \sum_{i=0}^{k-1} f_{i+1}/f_i = \sum_{i=0}^{k-1} (1 + f_{i-1}/f_i) = k + F_k. \end{aligned}$$

Corresponding to Lemma 7, we have

LEMMA 8. *$\sum_{\kappa=1}^{k} \bar{\phi}_{n+\kappa} \geqq F_k$, the equality holding only if $n = 0$.*

Proof. It follows from (7) that

$$(58) \qquad \bar{\phi}_{n+\kappa} = [0, q_{n+\kappa-1}, q_{n+\kappa-2}, \cdots, q_1].$$

If $n = 0$, then proceeding as in the proof of Lemma 6, we obtain $\bar{\phi}_{n+\kappa} = \bar{\phi}_\kappa = [0, 1, 1, \cdots, 1] = f_{\kappa-2}/f_{\kappa-1}$, $(\kappa = 1, 2, \cdots, k)$. Hence $\sum_{\kappa=1}^{k} \bar{\phi}_\kappa = \sum_{\kappa=1}^{k} f_{\kappa-2}/f_{\kappa-1} = \sum_{\kappa=0}^{k-1} f_{\kappa-1}/f_\kappa = F_k$.

Now consider the case $n \geqq 1$. We obtain from (7) $\bar{\phi}_{n+\kappa} = [0, 1, 1, \cdots, 1, q_n + \bar{\phi}_n]$, where the 1's occur $\kappa - 1$ times. Similarly to (57), we get

$$(59) \qquad \bar{\phi}_{n+\kappa} = [f_{\kappa-2}(q_n + \bar{\phi}_n) + f_{\kappa-3}]/[f_{\kappa-1}(q_n + \phi_n) + f_{\kappa-2}].$$

Upon differentiating with respect to $q_n + \phi_n$ and summing, we obtain

$$\sum_{\kappa=1}^{k} d\bar{\phi}_{n+\kappa}/d(q_n + \bar{\phi}_n) = \sum_{\kappa=1}^{k} (-1)^\kappa/[f_{\kappa-1}(q_n + \bar{\phi}_n) + f_{\kappa-2}]^2.$$

It follows as in Lemma 7 that $\sum_{\kappa=1}^{k} \bar{\phi}_{n+\kappa}$ is a decreasing function of $q_n + \bar{\phi}_n$.

Therefore, by (59),

$$\sum_{\kappa=1}^{k} \bar{\phi}_{n+\kappa} > \lim \sum_{\kappa=1}^{k} [f_{\kappa-2}(q_n + \bar{\phi}_n) + f_{\kappa-3}]/[f_{\kappa-1}(q_n + \bar{\phi}_n) + f_{\kappa-2}],$$

where $q_n + \bar{\phi}_n \rightarrow \infty$; i. e.,

$$\sum_{\kappa=1}^{k} \bar{\phi}_{n+\kappa} > \sum_{\kappa=0}^{k-1} f_{\kappa-1}/f_{\kappa} = F_k.$$

6. A generalization of a theorem of Hurwitz. We are now able to prove the following theorem:

THEOREM 17. *For every irrational number* ξ, *for every* $k > 0$, *and for every* n, *we have* $\sum_{\kappa=0}^{k-1} \lambda_{n+\kappa} > k + 2F_k$.

Proof. For $k = 1$, the theorem is trivial. For given n and $k > 1$, we define $\bar{\xi}$ as in (55). It follows from Lemma 5 that

$$\sum_{\kappa=0}^{k-1} \lambda_{n+\kappa} \geqq \sum_{\kappa=0}^{k-1} \bar{\lambda}_{n+\kappa}, \tag{60}$$

and from Lemmas 7 and 8 that

$$\sum_{\kappa=0}^{k-1} \bar{\lambda}_{n+\kappa} = \sum_{\kappa=1}^{k} (\bar{\xi}_{n+\kappa} + \bar{\phi}_{n+\kappa}) > k + 2F_k. \tag{61}$$

The theorem follows from (60) and (61).

Theorem 17 contains Theorems 4, 12, 13 and 15 as special cases. More exactly, it follows that the sum of any three consecutive λ_n is greater than 6 for every irrational number ξ. For the sums of four, five, six, seven and eight consecutive λ_n, we obtain correspondingly 25/3, 158/15, 767/60, 11711/780, and 31399/1820. Using this last bound for eight consecutive λ_n, we obtain $\underline{\lim}\, (\sum_{i=0}^{m-1} \lambda_i)/m > 2.1565$ instead of Theorem 16. However, this result can be improved as follows:

THEOREM 18. *For every irrational number* ξ, *we have*

$$\underline{\lim}\, (\sum_{i=0}^{m-1} \lambda_i)/m \geqq 5^{\frac{1}{2}}.$$

Proof. From Theorem 17, we have $\sum_{i=0}^{m-1} \lambda_i/m > (m + 2F_m)/m$. Now it is well known that the quotient of two consecutive Fibonacci numbers f_i/f_{i-1}, tends to $(5^{\frac{1}{2}} + 1)/2$ as i tends to infinity. Thus it follows from (46) that $\lim F_m/m = 2/(5^{\frac{1}{2}} + 1)$. Hence, $\underline{\lim}\, (\sum_{i=0}^{m-1} \lambda_i)/m \geqq 1 + 4/(5^{\frac{1}{2}} + 1) = 5^{\frac{1}{2}}$.

The example $\xi = (5^{\frac{1}{2}} + 1)/2$ shows that Theorem 18 cannot be improved. In addition, we wish to prove that the bound given in Theorem 17 is the best possible.

THEOREM 19. *For every given integer k and for every $\epsilon > 0$ it is possible to find irrational numbers ξ such that, for infinitely many values of n,*

$$\sum_{\kappa=0}^{k-1} \lambda_{n+\kappa} < k + 2F_k + \epsilon.$$

Proof. Let n be an arbitrary positive integer, and choose any number $\bar{\xi}$ for which $q_{n+\kappa} = 1$, $(\kappa = 1, 2, \cdot \cdot \cdot, k)$. Then it follows from (57) that

$$\sum_{\kappa=1}^{k} \bar{\xi}_{n+\kappa} = \sum_{i=0}^{k-1} \bar{\xi}_{n+k-i} = \sum_{i=0}^{k-1} (f_{i+1}\bar{\xi}_{n+k+1} + f_i)/(f_i\bar{\xi}_{n+k+1} + f_{i-1}).$$

It was shown in the proof of Lemma 7 that this sum is a decreasing function of $\bar{\xi}_{n+k+1}$ and that the right member tends to $k + F_k$ as $\bar{\xi}_{n+k+1}$ becomes infinite, i. e. as q_{n+k+1} becomes infinite. Thus, for sufficiently large q_{n+k+1}, we have

$$\sum_{\kappa=1}^{k} \bar{\xi}_{n+\kappa} < k + F_k + \tfrac{1}{2}\epsilon. \tag{62}$$

Similarly, it follows from the proof of Lemma 8 that $\lim\limits_{q_n \to \infty} \sum\limits_{\kappa=1}^{k} \bar{\phi}_{n+\kappa} < k + F_k + \frac{1}{2}\epsilon$. Hence, for sufficiently large q_n, we have

$$\sum_{\kappa=1}^{k} \bar{\phi}_{n+\kappa} < F_k + \tfrac{1}{2}\epsilon. \tag{63}$$

If we choose an irrational number ξ such that $q_{n+1} = q_{n+2} = \cdot \cdot \cdot = q_{n+k} = 1$ and such that q_n and q_{n+k+1} are so large that (62) and (63) are satisfied, then $\sum\limits_{\kappa=0}^{k-1} \lambda_{n+\kappa} < k + 2F_k + \epsilon$ for this number ξ.

If $n_1 < n_2 < \cdot \cdot \cdot$ is a set of positive integers such that $n_{i+1} - n_i \geqq k + 2$, $(i = 1, 2, \cdot \cdot \cdot)$, then in the same way we can construct irrational numbers ξ such that

$$\sum_{\kappa=0}^{k-1} \lambda_{n_i+\kappa} < k + 2F_k + \epsilon, \qquad (i = 1, 2, \cdot \cdot \cdot).$$

Thus the proof is complete.

We use the opportunity to remark that in the statement of Theorem 11, "$n > 2$" should be inserted after $\lambda_{n-2}, \lambda_{n-1}, \cdot \cdot \cdot, \lambda_{n+3m-1}$.

UNIVERSITY OF NORTH CAROLINA.

ON LINEAR FAMILIES OF INVOLUTIONS.*

By N. H. Kuiper.

O. Veblen and W. Givens [10, 11] studied the geometry of complex projective spaces. In this paper we follow their method and consider one of their topics: linear families of involutions. In Section 1 we mention the spin representation and we give a classification of (non-singular) linear families of involutions. In Section 2 we introduce an operation Ξ strongly connected with the spin theory. In Section 3, this is applied to the projective 7-dimensional space. Section 4 deals with a null system in a 6-dimensional quadratic hypersurface, Section 5 with the principle of triality, Section 6 deals with the 15-dimensional projective space. The theory reveals new aspects of the generalizations of the Kummer configurations (Section 7).

Much of what is presented can be carried over without difficulty to projective spaces over ground fields different from that of the complex numbers.

I wish to thank Professor Oswald Veblen for introducing me to the theory of spinors and for many valuable remarks concerning the content of this paper.

1. **Involutions.** We shall be concerned with a projective n-dimensional space P_n over the field of complex numbers, and with its linear subspaces $\{P_i\}$, shortly called spaces, of dimension $i = -1, 0, 1, \cdots, n$. After the choice of a coordinate system and homogeneous coordinates in P_n, a point ψ is represented by a class of $(n+1)$-tuples of numbers not all of which vanish ($\rho\psi^A$, $A = 1, \cdots, n+1$; $\rho \neq 0$ variable). The void set P_{-1} is represented by the $(n+1)$-tuple $(0, 0, \cdots, 0)$.

A collineation (γ) is a mapping of the family of subspaces of P_n into itself which preserves intersection relations, and which is determined by the fate of the points $(\psi \to \psi')$, expressed by equations of the form

$$\rho\psi'^A = \gamma^A{}_B\psi^B; \qquad \text{abbr.: } \rho\psi' = \gamma\psi \tag{1}$$

(the repeated index refers to summation over $B = 1, \cdots, n+1$). The matrices $\gamma^A{}_B$ and $\rho\gamma^A{}_B$ represent the same collineation ($\rho \neq 0$).

If the square of a collineation γ is the identity mapping, then γ is

---

* Received June 30, 1949; revised February 16, 1950.

called a non-singular involution and coordinates can be choosen such that the matrix of γ is

$$(2) \qquad \gamma = \begin{Vmatrix} \underline{1}_p & \\ & -\underline{1}_q \end{Vmatrix}$$

($\underline{1}_p$ is the unit matrix of p rows and p columns; zero in the empty places; $0 \leqq q, p \leqq n + 1 = p + q$). The points invariant under the involution are the points of a P_{p-1} and the points of a P_{q-1}. The involution is called a $(P_{p-1} - P_{q-1})$-reflection. It is the *identity* for $q = 0$ (or $p = 0$).

A collineation γ for which γ^2 transforms any point into P_{-1} is called a singular involution. For suitable coordinates,

$$(3) \qquad \gamma = \begin{Vmatrix} \underline{0} & \underline{1}_r \\ \underline{0} & \underline{0} \end{Vmatrix}, \quad 2r \leqq n + 1.$$

Let $\gamma_1, \cdots, \gamma_{m+1}$ be matrices representing involutions in P_n, and let the same be true for any linear combination $X^1\gamma_1 + X^2\gamma_2 + \cdots + X^{m+1}\gamma_{m+1} \equiv X^\alpha\gamma_\alpha$. Then the set of all these involutions is called a linear family. It follows that

$$(4) \qquad (X^\alpha\gamma_\alpha)^2 = (g_{\alpha\beta}X^\alpha X^\beta) \cdot \underline{1} \qquad (g_{\alpha\beta} = \text{const.}; \alpha, \beta = 1, \cdots, m + 1).$$

The involutions of the family are represented by points $(X^1, \cdots, X^{m+1})$ in a projective m-dimensional space R_m. In particular the singular involutions are represented on the quadric $g_{\alpha\beta}X^\alpha X^\beta = 0$. We shall restrict our considerations to non-singular families of involutions, those for which this quadric is not degenerate. A coordinate transformation in R_m corresponds to a change of base for the involutions of the family. For suitable coordinates in R_m we have $g_{\alpha\beta} = 0$ for $\alpha \neq \beta$ and $g_{\alpha\alpha} = 1$. Then the following form is obtained:

$$(5) \qquad (X^\alpha\gamma_\alpha)^2 = (\Sigma_\alpha X^\alpha X^\alpha) \cdot \underline{1}; \qquad \gamma_\alpha\gamma_\beta + \gamma_\beta\gamma_\alpha = 2\delta_{\alpha\beta} \cdot \underline{1}.$$

The space P_n assigned with a non-singular linear family of involutions $L(m + 1, n + 1)$ is called spin space. Any collineation in P_n which leaves this family of involutions invariant is called a spin transformation. The spin transformations form a group G_s.

The related space R_m is called motion space. A collineation in R_m which leaves the quadric invariant is called a motion. The motions form a group G_m.

It is easily seen that any spin transformation interchanges the involutions of the family in a way which is determined by a motion in R_m, and any motion in R_m can be obtained in this way. Consequently there is a homomorphism of G_s onto G_m. If this is an isomorphism, then G_s is called the spin representation of G_m.[1] Two non-singular involutions γ_1, γ_2 are called *anti-commuting* if ($\underline{a}$) they belong to some pencil of involutions, and ($\underline{b}$) they are represented in the related motion space R_1 by points polar with respect to the quadric in R_1. Conditions ($\underline{a}$) and ($\underline{b}$) are equivalent to the matrix condition

$$\gamma_1\gamma_2 + \gamma_2\gamma_1 = 0. \tag{6}$$

The following normal form is obtained with suitable coordinates in P_n:

$$\gamma_1 = \begin{Vmatrix} & \underline{1} \\ \underline{1} & \end{Vmatrix}, \qquad \gamma_2 = -i \begin{Vmatrix} & \underline{1} \\ -\underline{1} & \end{Vmatrix}; \qquad \text{pencil: } X^1\gamma_1 + X^2\gamma_2. \tag{7}$$

This pencil of involutions introduces another pencil in the line which is the join of the points $(a_1, \cdots, a_r, 0, \cdots, 0)$ and $(0, \cdots, 0, a_1 \cdots, a_r)$, where $2r = n + 1$. The set of such lines is called line regulus. Any line of the line regulus meets any axis $(= P_{r-1})$ of the axis regulus consisting of the one-parameter set of pointwise invariant spaces of the involutions of the pencil. The line and axis regulus cover the same r-dimensional manifold in P_n. If we define an involution of axes on an axis regulus in an obvious way, there results the following geometrical description of a pencil of involutions:

THEOREM 1. *The pairs of invariant spaces of the involutions of a non-singular pencil $L(2, n+1)$ in P_n are the pairs of conjugate axes in an involution of axes on an axis regulus (n is odd by necessity).*

Any non-singular linear family of involutions $L(m+1, n+1)$ $(m \geqq 1)$ admits, for suitable coördinates in P_n and in R_m, the following description (compare (5) and (7)):

$$X^i \begin{Vmatrix} \gamma_i & \\ & -\gamma_i \end{Vmatrix} + X^m \begin{Vmatrix} & \underline{1} \\ \underline{1} & \end{Vmatrix} + X^{m+1} i \begin{Vmatrix} & \underline{1} \\ -\underline{1} & \end{Vmatrix}, \tag{8}$$

$$\gamma_i^2 = \underline{1}; \qquad \gamma_i\gamma_j + \gamma_j\gamma_i = 0\,(i \neq j); \qquad i, j = 1, \cdots, m-1.$$

---

[1] There is a difference between this theory and the theory which is usually called spin theory, in which a Euclidean centered space E_{m+1} occurs instead of R_m, and an affine centered space A_{n+1} instead of P_n. One first introduces the Euclidean space. Much of our theory can be carried over to this theory in an obvious way. We also remark that we do not enter into interesting problems which occur under restriction of the ground field to that of the real numbers. Cf. references [7, 8, 10, 11, 12, 13, 14, 19].

The last two terms in (8) determine a pencil. Because $X^i\gamma_i$ represents an $L(m-1,(n+1)/2)$, we now easily obtain, by induction on r, the following theorems:

THEOREM 2a. *Non-singular linear families of involutions of the type* $L(s, 2^r \cdot p), s > 2r+1$, *$p$ odd, do not exist.*

THEOREM 2b. *Any two families of the type* $L(2s, 2^r \cdot p), s \leqq r$, *are projectively equal.*

THEOREM 2c. *The classes of projectively equal families of the type* $L(2s+1, 2^r \cdot p)$, $s \leqq r$, *are in one-to-one correspondence with the classes of projectively equal* $L(1, 2^{r-s} \cdot p)$ *which are involutions in* $P(2^{r-s} \cdot p - 1)$. *The classification of involutions follows from* (2).

(8) contains the key for the construction of an $L(m+1, 2r)$ from an $L(m-1, r)$. In particular, considering that $L(1,1)$ is the identity transformation in P_0, described by the one-rowed matrix $\gamma_i = \| 1 \|$, $i = 1$, we obtain from (8) a normal expression for

$$(9) \qquad \begin{cases} L(3,2): \quad X^1 \begin{Vmatrix} 1 & \\ & -1 \end{Vmatrix} + X^2 \begin{Vmatrix} & 1 \\ 1 & \end{Vmatrix} + X^3 \cdot i \begin{Vmatrix} & 1 \\ -\underline{1} & \end{Vmatrix} \\ L(5,4),\ L(2r+1, 2r) \end{cases}$$

(cf. [10], pp. 5-6). For the cases (9) it is known that the spin group G_s is isomorphic to the motion group G_m.

2. The operation Ξ. Let a spin space be given by a non-singular linear family of involutions $L(m+1, n+1)$ in P_n. To every point $\psi \varepsilon P_n$ we assign a conjugate space $\Xi\psi$, which is the union, and also the join, of all images of ψ under all involutions of $L(m+1, n+1)$. This operation Ξ is invariant under spin transformations, because so is $L(m+1, n+1)$. We shall examine some of these operations which are determined by (maximal) families $L(m+1, n+1) = L(2r+1, 2^r)$.

$L(3,2)$, the first interesting case, consists of all involutions in P_1, except $\gamma = \underline{0}(9)$. Ξ transforms every point ψ onto P_1. Corresponding to every point ψ in P_1, there exists one (singular) involution which transforms ψ onto P_{-1}, and a pencil of involutions which transform ψ into ψ. *Thus points of the spin space* (P_1) *are* 1-1 *represented by points of the invariant quadric* (= *conic section*) *in the motion space* R_2, *and also* 1-1 *by the tangents to this quadric.* The spin group is obviously the group of all non-singular projective transformations in P_1, isomorphic with the motion group in R_2.

As to $L(5,4)$, it is seen from (9) that the images under the involutions of $L(5,4)$ of a point with coordinates (a,b,c,d) are the points (x,y,z,u) for which

$$dx - au + bz - cy = 0. \tag{10}$$

Since this is the equation of a null system, we obtain

THEOREM 3a. *The operation* Ξ *determined by a family of involutions* $L(5,4)$ *in* P_3 *is a null system.*

Any line in P_3 has a conjugate line with respect to the null system Ξ (cf. [7]). The self-conjugate lines form a linear line complex. Two pairs of conjugate lines are four lines on a quadric, hence they can be considered as the invariant axes of two involutions of a pencil of involutions (Theorem 1). The family of involutions the invariant spaces of which are the pairs of conjugate lines in the null system, is then a linear family which coincides with $L(5,4)$.

THEOREM 3b. *The pairs of invariant axes of an* $L(5,4)$ *in* P_3 *are the pairs of lines conjugate with respect to a null system* Ξ *in* P_3.

COROLLARY. *The spin group with respect to an* $L(5,4)$ *is the group of projective transformations which leave a null system invariant* (*in* P_3); *it is isomorphic with the group of motions in* R_4.

$L(5,4)$ is a four-dimensional family R_4. Ξ transforms a point into a plane. The linearity of involutions as well as of the family then implies the following fact: Corresponding to every point $\psi \varepsilon P_3$, there exists exactly one pencil (R_1) of singular involutions in $L(5,4)$ which transform ψ into P_{-1}, and also one R_2 of involutions all of which transform ψ into ψ. *Hence points of spin space* P_3 *are* 1-1 *represented by lines of the three-dimensional invariant quadric in motion space* R_4, *and also* 1-1 *by planes* (R_2) *which are tangent to this quadric along such lines.*

3. The operation Ξ in P_7. The complete description of Ξ determined by a family $L(7,8)$ in P_7 will be given as a summary of the following lemmas.

LEMMA 3.1. *If* $\psi \varepsilon$ *the invariant space of a singular involution of* $L(7,8)$, *then* $\Xi\psi$ *is a* P_3 *and* $\psi \varepsilon \Xi\psi$.

Proof. We choose coordinates in motion space and in spin space such that $L(7,8)$ has the expression (9) and that the given singular involution is

$$\tfrac{1}{2}(\gamma_6 - i\gamma_7) = \tfrac{1}{2}\left(\begin{Vmatrix} & \underline{1} \\ \underline{1} & \end{Vmatrix} - i \cdot i \begin{Vmatrix} & \underline{1} \\ -\underline{1} & \end{Vmatrix} \right) = \begin{Vmatrix} & \underline{1} \\ \underline{0} & \end{Vmatrix}.$$

The invariant space a^0 of this involution consists of the points $(a, b, c, d, 0, 0, 0, 0)$ $((\gamma_6 - i\gamma_7)\psi = 0)$. The involutions γ_j, $j = 1, \cdots, 5$, leave a^0 invariant, and they introduce an $L(5,4)$, $X^j\gamma_j$, in a^0. Hence if $\psi = (x^1, x^2, x^3, x^4, 0, 0, 0, 0)$, then $\Xi\psi$ is the join of a plane which contains ψ and is obtained from the null system introduced by that $L(5,4)$ in a^0 (Theorem 3a) *and* the point $\tfrac{1}{2}(\gamma_6 + i \cdot \gamma_7)\psi = (0, 0, 0, 0, x^1, x^2, x^3, x^4)$; q. e. d.

The plane just mentioned consists of the points $(y^1, y^2, y^3, y^4, 0, 0, 0, 0)$ for which

$$x^4y^1 - x^1y^4 + x^2y^3 - x^3y^2 = 0 \tag{11}$$

(cf. (10)). Then $\Xi\psi$ is the set of points $\lambda y^1, \lambda y^2, \lambda y^3, \lambda y^4, \mu x^1, \mu x^2, \mu x^3, \mu x^4)$ for which (11) holds. This implies the following

LEMMA 3.2. *The union of the points $X^\alpha\gamma_\alpha\psi$, ψ variable in a^0, is a quadric Q_6 in P_7, consisting of the points $(y^1, y^2, y^3, y^4, x^1, x^2, x^3, x^4)$ satisfying* (11).

LEMMA 3.3. *If ϕ is a point in Q_6, then $\Xi\phi$ is a P_3 and $\phi \varepsilon \Xi\phi$.*

Proof. ϕ is the image under some involution of the family of some point ψ in a^0. The 6-dimensional family of involutions transforms ψ into the points of a P_3. This implies the existence of an R_2 and of an R_3 of involutions in the family all of which transform ψ into P_{-1} and ϕ, respectively. The non-singular involutions of the R_3 also transform ϕ into ψ, and a suitable linear combination of two of these transforms ϕ into P_{-1}. Therefore ϕ is contained in the invariant space of some singular involution of the family, hence Lemma 3.1 can be applied.

LEMMA 3.4 (cf. [17], p. 29). *Q_6 in P_7 contains two families of axes (P_3's), axes^1 and axes^2. The axes of the same (different) family intersect in a P_3, P_1 or P_{-1} (a P_2 or P_0).*

LEMMA 3.5. *The invariant axes of the involutions of $L(7,8)$ are axes of one family (say axes^1) in the Q_6 of Lemma 3.2.*

Proof. $\gamma_6 - i\gamma_7$ determines with an involution $\gamma = X^{1\alpha}\gamma_\alpha$ which does not anti-commute with $\gamma_6 - i\gamma_7$ a non-singular pencil of involutions (γ and $\gamma_6 - i\gamma_7$ are in motion space represented by points not polar with respect to the invariant quadric in R_6). We may choose coordinates such that this pencil is $X^6\gamma_6 + X^7\gamma_7$ and $L(7,8)$ has the expression (9). Among the images of the points of a^0 under the involutions of the family, we find all points of the line regulus and axis regulus determined by this pencil of involutions (Theorem 1). In particular we find the points of the invariant spaces of γ, which are therefore contained in Q_6. Any involution of the family which *does* anti-commute with $\gamma_6 - i\gamma_7$ is contained in some pencil of involutions all but one of which have invariant space, now known to be contained in Q_6. Hence the invariant axes of any involution of the family are in Q_6. The invariant axes of the involutions of a non-singular pencil are mutually disjoint and therefore (Lemma 3.4) the invariant axes are of one family in Q_6.

LEMMA 3.6. *If $\psi \not\in Q_6$, then $\Xi\psi$ is the polar hyperplane of ψ with respect to Q_6.*

Proof. ψ is not contained in the invariant space of any singular involution of $L(7,8)$. The images of ψ under the involutions $\gamma_1, \cdots, \gamma_7$ are then seven independent points in P_7, the join of which is a P_6. The points ψ and $\gamma_a\psi$ $(\alpha = 1, \cdot, \text{or } 7)$ span a line, invariant under γ_a, in which γ_a operates as an (ordinary) involution. Hence ψ and $\gamma_a\psi$ lie harmonic with respect to two points in the invariant spaces of γ_a, that is, in Q_6. Hence Lemma 3.6 follows.

LEMMA 3.7. *If $\psi \varepsilon Q_6$, then $\Xi\psi$ is an axis^2 in Q_6.*

Proof. As in the proof of Lemma 3.3, we find that any point of $\Xi\psi$ is contained in the invariant space of some singular involution of the family and in Q_6. Hence $\Xi\psi$ is an axis in Q_6. For the special choice $\psi \varepsilon a^0$, which is no restriction for the proof, we see that $\Xi\psi$ and the axis^1 a^0 intersect in a plane. This and Lemma 3.4 imply that $\Xi\psi$ is an axis^2.

The correspondence $\psi \rightarrow \Xi\psi$, $\psi \varepsilon Q_6$, can be extended to a 1-1 correspondence of linear spaces in Q_6 for which the name null system in Q_6 will turn out to be appropriate. We summarize the lemmas in

THEOREM 4. *The operation $\psi \rightarrow \Xi\psi$, determined by a family of involutions $L(7,8)$ in P_7, coincides with a polarity with respect to a quadric Q_6 for those points which are not in Q_6; it coincides with a null system in Q_6 for those points which are in Q_6.*

4. The null system Ξ in Q_6.

LEMMA 4. 1. *If ψ, ϕ are points in Q_6, and $\phi \varepsilon \Xi\psi$, then $\psi \varepsilon \Xi\phi$.*

It follows, that if l is a line in Q_6, and if $\psi \varepsilon l \subset \Xi\psi$ holds for some point ψ, then the same is true of any point of l. The line l is then self-conjugate with respect to the null system Ξ. The family of all self-conjugate lines is called a linear line complex in Q_6. Obviously the linear line complex in Q_6 determines the null system.

THEOREM 5. *A null system and a linear complex in a Q_6 are determined by an involution of the axes of an axis regulus in that Q_6.*

The figure consisting of the involution of axes^1 which determines the pencil $X^6\gamma_6 + X^7\gamma_7$ in $L(7, 8)$ *and* the quadric Q_6 (of Section 3) has no projective invariants. (Any two such figures are projectively equal.) From the properties of $L(7, 8)$ we now obtain the geometrical *construction of the null system Ξ and the line complex,* and hence the proof of Theorem 5, as follows: Fix an involution of axes^1 in an axis^1 regulus in a Q_6. With respect to this involution, three kinds of points (ψ) occur in Q_6, the conjugate spaces of which under the null system are constructed as follows:

(a) ψ is contained in a double axis^1 of the axis^1 involution (say $\psi \varepsilon$ the invariant space a^0 of $\gamma_6 - i\gamma_7$). Then $\Xi\psi$ is the unique axis^2 which contains the regulus line through ψ (the join $\psi \cup \gamma_6\psi$), and which intersects a^0 in a P_2 $(=\psi \cup \gamma_1\psi \cup \gamma_2\psi \cup \gamma_3\psi \cup \gamma_4\psi \cup \gamma_5\psi$; the join).

(b) ψ is contained in another axis^1 of the axis^1 regulus (say in γ_6^+, one of the invariant axes of γ_6). Then $\Xi\psi$ is the unique axis^2 which contains ψ and intersects the conjugate axis^1 (γ_6^-, the other invariant axis of γ_6) in a P_2 $(=\gamma_1\psi \cup \gamma_2\psi \cup \cdots \cup \gamma_5\psi)$.

(c) In the remaining case, let m_1 and m_2 be two lines which contain ψ and each of which intersects two conjugate axis^1 of the axes^1 involution ($m_1 : \psi \cup \gamma_6\psi$; $m_2 : \psi \cup \gamma_7\psi$). Then $\Xi\psi$ is the unique axis^2, which contains the plane of these two lines (plane: $\psi \cup \gamma_6\psi \cup \gamma_7\psi$).

From now on in this section we shall consider a fixed $L(7, 8)$ in the normal form (9). $X^6\gamma_6 + X^7\gamma_7$ is *the* pencil, belonging to *the* line regulus of lines m and to *the* axis^1 regulus, both with *point set M.* It is our aim to extend the null system to a transformation (also called null system) which operates on all points, lines and axes in Q_6.

Corresponding to any axis^2, a^2, we construct a "conjugate" point Ξa_2, of which a^2 is the conjugate. This proves that the mapping $\psi \rightarrow \Xi\psi$, $\psi \varepsilon Q_6$, is a 1-1 mapping of the family of points onto the family of axes^2 in Q_3. First suppose that a^2 intersects one of the axes^1, say a^1, of *the* axis^1 regulus, in a plane (see Lemma 3.4). Then it intersects the others in points. If ψ is one of these points, not in a^1, then a^2 is the unique axis^2 which contains ψ and intersects a^1 in a plane. But then a^2 intersects the other axes^1 of the axis^1 regulus in the points of a regulus line m. And Ξa^2 is the intersection point of m and the axis^1 which is conjugate in the axis^1 involution to a^1, in accordance with (a) and (b) above.

Next, suppose that a^2 intersects all axes^1 of the regulus in points, among others in the points ψ_1 and ψ_2 on the regulus lines m_1 and m_2, respectively. If m_1 equals m_2, we have the case dealt with above. Otherwise the $P_3: m_1 \cup m_2$ intersects M in a two-dimensional quadric Q_2 ($=$ a line regulus), and also in the line $\psi_1 \cup \psi_2$. This is only possible if $m_1 \cup m_2$ is completely contained in Q_6. It intersects the axes^1 of the regulus in lines and is therefore itself an axis^1, say b^1. Then b^1 intersects a^2 among others in the line $\psi_1 \cup \psi_2$, hence in a plane (Lemma 3.4). This plane intersects $Q_2 = b^1 \cap M$ in a conic section. The latter is intersected by the axis^1 involution in a point involution. The joins of corresponding points of this involution are lines of a pencil of lines, the center of which is the point Ξa^2, in accordance with (c).

So far we proved one part of the next theorem, for the formulation of which we need a definition of adjacence: A line and a point are called adjacent if they meet. A line and an axis (in Q_6) are adjacent if the line is contained in the axis. Two other elements of the family of points, lines and axes in Q_6 are adjacent if a third element exists, adjacent with both in the sense defined *so far*. For example, a point and an axis are adjacent if the point is in the axis. Two axes^1 are adjacent if their intersection is a line, etc.

THEOREM 6. *The null system Ξ which operates on points in Q_6 can be extended to a "null system," Ξ, which operates on points, lines axes^1 and axes^2 in Q_6 in the following manner:*

1) *If ψ is a point, then $\Xi\Xi\psi = \psi$.*
2) *If a^2 is an axis^2, then Ξa^2 is a point and $\Xi\Xi a^2 = a^2$.*
3) *If l is a line, so is $\bar{l} = \Xi l$, and $\Xi\Xi l = l$.*
4) *If a^1 is an axis^1, so is $\bar{a}^1 = \Xi a^1$, and $\Xi\Xi a^1 = a^1$.*
5) *The null system preserves the relation of adjacence.*

COROLLARY. *There exists a* 1-1 *rational representation of the axes (say axes^2) of one family on a Q_6, onto the points of Q_6, namely, Ξ.*

Cf. [15]. Intersection with a general P_6 in P_7, which intersects Q_6 in a Q_5, yields a representation of the planes in a Q_5 in terms of the points of a Q_6. Cf. [16], p. 80.

Proof of Theorem 6. The self-conjugate lines were considered immediately after Lemma 4.1. Suppose that l is not a self-conjugate line. Consider two axes^2 which contain l, say a^2 and b^2, with conjugates the points $\psi = \Xi a^2$ and $\phi = \Xi b^2$. If the point $\chi \varepsilon l$, then $\chi \cup \psi$ and $\chi \cup \phi$ are lines of the line complex. Hence the same is true of all lines of the pencil determined by these two lines. Since χ was arbitrary on l, it follows that all lines which meet l and $\bar{l} = \psi \cup \phi$ belong to the linear line complex. The relation between line complex and null system then implies Theorem 6.3 (and part of 6.5).

Next we consider an axis^1, a^1. Any plane in a^1 is contained in exactly one axis^2. The planes of a^1 which contain a line l are contained in axis^2 the conjugates of which are the points of $\bar{l}$. The planes of a^1 which contain at least one edge of a tetrahedron in a^1 are therefore contained in axes^2 the conjugates of which are the points of the edges of another tetrahedron in Q_6. The join of the points of this tetrahedron is an axis in Q_6. Hence, by linearity, the planes of a^1 are contained in axes^2 the conjugates of which are the points of the axis just found. This is not an axis^2, hence an axis^1; so that Ξa^1 is the conjugate, $\bar{a}^1$. Obviously the conjugates of the lines of a^1 are the lines of $\bar{a}^1$. The preservation of adjacence can now be checked for all cases.

The construction of the null system shows that the conjugate of an axis^1 of the axis^1 involution in *the* axis^1 regulus is the conjugate with respect to the null system as well as with respect to the axis^1 involution. The invariant spaces of an involution of *the* pencil $X^6\gamma_6 + X^7\gamma_7$ are conjugate axes^1 of the null system in Q_6. Because the operation Ξ does not distinguish between the non-singular pencils of involutions of $L(7, 8)$, the same is true of the invariant spaces of any involution in $L(7, 8)$.

The converse can be seen as follows: Every line in Q_6 which intersects two different conjugate axes^1 belongs to the linear line complex in Q_6, and every line of this complex which meets an axis^1 also meets its conjugate axis^1. Taking into consideration Theorem 1, it follows that the family of all involutions, having as a pair of invariant spaces a pair of (different) conjugate axes^1 suitably extended with singular involutions, is a linear family; the latter not only contains the given $L(7, 8)$ but coincides with it since it has the same dimension.

THEOREM 7. *The pairs of (possibly identical) invariant spaces of the involutions of a linear family* $L(7,8)$ *in* P_7 *are the pairs of axes*^{1} *conjugate with respect to a null system in a quadric* Q_6*, and vice versa.*

Remarks. Only points in spin space $P_7(L(7,8))$ which are in the Q_6, admit an interesting representation in motion space R_6, in terms of planes in the invariant quadric $Q_5 \subset R_6$ (cf. the end of Section **2**). The spin group consists here of all projective transformations in P_7 which leave quadric and null system invariant, isomorphic with the motion group in R_6.

5. The principle of triality. Consider the family F of linear subspaces, of a dimension number different from 2, of a $Q_6 \subset P_7$: the points (symbol 0), the lines, the axes^1 (symbol 1), and the axes^2 (symbol 2). The triality group $\mathfrak{G}$ of Q_6 is by definition the group of 1-1 mappings of F onto itself, generated by (1) the group G of those collineations in P_7, which leave Q_6 and each family of axes in Q_6 invariant (proper collineations in Q_6); (2) one collineation in P_7 which interchanges axes^1 and axes^2, e. g., a point polar hyperplane reflection Λ in P_7; (3) one null system Ξ in Q_6. All elements of $\mathfrak{G}$ transform lines into lines. $\mathfrak{G}$, considered as a topological space, has the following six components, the elements of which have the indicated activity:

Component		Name of elements	Activity
I:	G	Proper collineations	$0 \rightarrow 0, 1 \rightarrow 1, 2 \rightarrow 2$
II:	$G\Lambda$	Improper collineations	$0 \rightarrow 0, 1 \rightleftarrows 2$
III:	$G\Xi$	^{1}correlations	$1 \rightarrow 1, 0 \rightleftarrows 2$
IV:	$G\Lambda\Xi\Lambda$	^{2}correlations	$2 \rightarrow 2, 0 \rightleftarrows 1$
V:	$G\Lambda\Xi$	^{1}trialities	$0 \rightarrow 1 \rightarrow 2 \rightarrow 0$
VI:	$G\Xi\Lambda$	^{2}trialities	$0 \leftarrow 1 \leftarrow 2 \leftarrow 0$

The extended triality group in Q_6 is generated by $\mathfrak{G}$ and one anti-involution in P_7 which leaves the quadric invariant. If the coefficients of the equation which represents Q_6 are real (as above), then such an anti-involution is obtained from the automorphism $i \rightarrow -i$ of the ground field. The anti-involution then consists in the mapping of any point onto the point with complex conjugate coordinates. This makes obvious the following fact (cf. [1, 2]):

THEOREM 8. *The transformations of the (extended) triality group preserve the relations of adjacence between points, lines and axes in* Q_6.

This theorem is the expression for a principle of triality analogous to the principle of duality (point-hyperplane, etc.) in the geometry of the projective space P_n. By the principle of triality, theorems on points, lines, and axes in Q_6 can be carried over into other theorems. Consider, for example, the following theorem (cf. [18], p. 51): A 1-1 point transformation of Q_6, which preserves adjacence of points, is a proper or improper collineation or anti-collineation of Q_6 (if Q_6 is represented by an equation with real coefficients, then the "anti" refers to the automorphism $i \to -i$). The triality principle transforms this theorem into

THEOREM 9. *A 1-1 adjacence-preserving transformation of the axes*[1] *in a* Q_6 *is a proper collineation, a proper anti-collineation, a* ^{1}*correlation or an anti-*^{1}*correlation* (*an example is the null system*).[2]

Another application of the principle of triality results as follows: Under a ^{2}correlation the improper collineations in Q_6 are interchanged with the ^{1}correlations; in particular the point-hyperplane reflections of Q_6 (in P_7) are interchanged with the ^{1}null systems (like Ξ). The ^{1}null systems as well as the $L(7,8)$'s with a fixed Q_6 in P_7, are therefore rationally representable (1-1) by the points of a P_7 not in a Q_6 in that P_7. The points of the Q_6 correspond to degenerated null systems.

6. The operation Ξ in P_{15}. In this section Ξ will indicate the operation which assigns, to every point $\psi \in P_{15}$, the linear space which is the join of the images of ψ under the involutions of an $L(9,16)$. The analysis of Ξ is analogous to that in preceding cases. Let $L(9,16)$ have the normal form (9). Then $X^8\gamma_8 + X^9\gamma_9$ is *the* pencil of involutions, the invariant spaces of which form *the* axis regulus (P_7's) with M as point set. M also covers *the* line regulus (Theorem 1). The invariant spaces of $\gamma_8 - i\gamma_9$ and $\gamma_8 + i\gamma_9$

---

[2] *Remarks*: Wei Liang Chow ([18] p. 55) proved that an adjacence-preserving 1-1 mapping of the axes (P_r) of one family in a quadric Q_{2r} is a collineation or an anti-collineation if $r > 3$. He drew my attention to the interesting exceptional case $r = 3$, of which Theorem 9 gives a specific description.

The principle of triality originates from E. Study [6]. In his work he gives an interesting interpretation of the principle of triality in terms of point pairs, proper motions and improper motions in an elliptic three-dimensional space.

Many applications of the principle are found in the book of Weiss ([16], p. 154 ff.).

E. Cartan [7] arrived at the principle of triality in group theory. Considering Lie groups and their automorphisms, he has shown that the group G^* of collineations and anti-collineations in Q_6 is exceptional. All its automorphisms are obtained from the inner automorphisms $(\mathfrak{g} \to \mathfrak{h}^{-1} \cdot \mathfrak{g} \cdot \mathfrak{h})$ of the extended triality group. The factor group of the group of automorphisms over the group of inner automorphisms of G^* is of order 6.

are a^0 and b^0, respectively. $X^i\gamma_i$, $i = 1, \cdots, 7$, introduces an $L(7, 8)$ in the P_7 a^0. The invariant spaces of these involutions in a^0 lie in the Q_6 of points $(x^1, x^2, x^3, x^4, y^1, y^2, y^3, y^4, 0, \cdots, 0)$ obeying (11) of Section **3**.

If ψ $(x^1, \cdots, y^4, 0, \cdots, 0)$ is in a^0 but not in Q_6, then $\Xi\psi$ is a P_7 which does not contain ψ and which is the join of the polar-P_6 of ψ with respect to the Q_6 in a^0, and the point $\gamma_8\psi = -\gamma_9\psi$ $(0, \cdots, 0, x^1, \cdots, y^4)$ in b^0. For ψ variable (in a^0 but not in Q_6), the spaces $\Xi\psi$ cover the set of points $(u^1, u^2, u^3, u^4, v^1, \cdots, v^4, x^1, \cdots, x^4, y^1, \cdots, y^4)$

$$x^4v^1 + u^4y^1 - x^1v^4 - u^1y^4 + x^2v^3 + u^2y^3 - x^3v^2 - u^3y^2 = 0, \tag{12}$$

$$x^4y^1 - x^1y^4 + x^2y^3 - x^3y^2 \neq 0 \tag{13}$$

(cf. (11)). Any point of this set is contained in the invariant space of some singular involution of $L(9, 16)$ (cf. the end of Section **2**). Conversely, if there is given in the family some singular involution which does not anti-commute with $\gamma_8 - i\gamma_9$, then its invariant space is contained in the Q_{14} of (12). This is seen from the example $\gamma_8 + i\gamma_9$. The locus of points contained in the invariant space of some singular involution of $L(9, 16)$ is an algebraic manifold, hence it is (12). If two singular involutions do not anti-commute, then they are two involutions of a non-singular pencil, and their invariant spaces are disjoint (cf. the representing points in motion space R_8). In consequence of the generalization of Lemma 3. 4, these invariant spaces are axes of one family in the Q_{14}, (12). Hence the invariant axes of the singular involutions of $L(9, 16)$ are (say) axes^1 in Q_{14}.

If ψ is in a^0 but not in Q_6 (in a^0), then $\Xi\psi$ is a P_7 in Q_{14} and intersects a^0 in a P_6. Then $\Xi\psi$ is an axis^2 of Q_{14}. The same is true of any point $\psi \varepsilon Q_{14}$ for which $\Xi\psi$ is a P_7. All these points lie in the invariant space of exactly one singular involution of the family. Because $\psi \not\varepsilon \Xi\psi$, these points are not contained in the invariant space of any non-singular involution of $L(9, 16)$.

Because every point in Q_{14} lies in the invariant space of some singular involution, all kinds of points can be found in a^0. So we still have to consider the points in the Q_6 in a^0. If ϕ is such a point, then $\Xi\phi$ is the P_4 spanned by the $P_3 : \phi \cup \gamma_1\phi \cup \cdots \cup \gamma_7\phi$ and the point $\gamma_8\phi$ in b^0. The locus of all points like ϕ in Q_{14} has at the point $\gamma_8\phi$ a tangent space, which contains the P_4 just mentioned and also tangent P_6 to *the* Q_6 in b^0. This locus is therefore of dimension $\geqq 10$. It is the intersection of the cone-like manifolds C_{14},[3] which are constructed as follows: Take any two not anti-commuting singular involutions of $L(9,16)$, e.g., $\gamma_8 - i\gamma_9$ and $\gamma_8 + i\gamma_9$ with invariant spaces a^0 and b^0. The locus of the joins of one point in b^0 and one point in the invariant Q_6 in a^0 is a C_{14} as required:

---

[3] T. G. Room has pointed out that this locus is an intersection of quadratic hypersurfaces C_{14}.

$$(u^1, \cdots, v^4, x^1, \cdots, y^4) \qquad x^4y^1 - x^1y^4 - x^2y^3 - x^3y^2 = 0.$$

The intersection of all C_{14}'s is easily seen to be of dimension $\leqq 10$, hence of dimension $= 10$. We call it T_{10}.

The algebraic manifold which consists of all points of P_{15} contained in the invariant space of some involution of $L(9,16)$, contains Q_{14} but also *the* axis regulus of $X^8\gamma_8 + X^9\gamma_9$. This locus is therefore of dimension > 14, and must be P_{15} itself.

Apart from some details we have now proved

THEOREM 10. *A non-singular family of involutions* $L = L(9,16)$ *in* P_{15} *determines an invariant quadric* Q_{14}, *and an invariant manifold* $T_{10} \subset Q_{14}$ *which is an intersection of cones* C_{14}. *Every invariant axis of every involution of* L *meets* T_{10} *in a* Q_6. *Every point* ψ *in* P_{15}, *but not in* T_{10}, *is contained in an invariant axis of exactly one involution of* L, *which is singular* (*non-singular*) *if* ψ *is contained* (*not contained*) *in* Q_{14}. *The invariant axes of the singular involutions of* L *are axes of one family in* Q_4 (*say axes*^{1}).

The operation Ξ *determined by* $L(9,16)$ *is as follows*: (1) *If* $\psi \notin Q_{14}$, *then* $\Xi\psi$ *is a* P_8 *which contains* ψ *and intersects* T_{10} *in a* Q_6. (2) *If* $\psi \in Q_{14}$, $\psi \notin T_{10}$, *then* $\Xi\psi$ *is an axis*^{2} *in* Q_{14} *which does not contain* ψ *and intersects* T_{10} *in a* Q_5. (3) *If* $\psi \in T_{10}$, *then* $\Xi\psi$ *is a* P_4 *in* T_{10}, *and* $\psi \in \Xi\psi$.

Concerning the general case of an operation Ξ determined by a linear family of involutions $L(2r+1, 2^r)$ in P_{2^r-1}, we only mention the following property: Let ψ be a point in this space and let $r - \lambda(\psi)$ be the maximal number of linearly independent, singular involutions of the family with invariant spaces which contain ψ. Then either $\lambda = 0$ or $3 \leqq \lambda \leqq r$. If $\lambda = 0, 4$, then $\psi \in \Xi\psi$. If $\lambda = 3$, then $\psi \notin \Xi\psi$. Otherwise both cases may occur. If $\lambda = 0$, then the point ψ in spin space is represented (1-1) by an R_{r-1} in motion space R_{2r}.

O. Veblen and W. Givens [10] has shown that an $L(2r+1, 2^r)$ determines an invariant involutory correlation C, which is a polarity in a quadric if $r \equiv 0, 3 \pmod 4$ and a null system (in the ordinary sense) if $r \equiv 1, 2 \pmod 4$. If $r = 1, 2$, then C coincides with our operation Ξ. If $r = 3, 4$, then C is the polarity with respect to the invariant quadric (Q_6, Q_{14}) which we mentioned. If $r = 3$, then C coincides nearly with Ξ. For $r \geqq 4$, C and Ξ are no longer related in such a simple way.

7. **Kummer configurations.** It was suggested by T. G. Room that there might exist connections between the subject of this paper and the configuration $Cf(16_6)$ of Kummer [1, 3, 4], generalized by Wirtinger [2], Barrau [5] and Room [9]. We shall show that such connections exist, and that the operation Ξ reveals some new aspects of those configurations.

The basic figure is a set of $2r+1$ anti-commuting involutions $\gamma_1, \cdots, \gamma_{2r+1}$ in P and the invariant involutory correlation C in P_{2^r-1} (cf. above). The involutions are the base of a linear family $L(2r+1, 2^r)$ which again determines an operation Ξ. The involutions $\gamma_1, \cdots, \gamma_{2r+1}$ and the correlation C generate a group F consisting of 2^{2r} involutions and 2^{2r} involutoric correlations of which $(2^r-1)2^{r-1}$ are null systems and $(2^r+1)2^{r-1}$ are polarities in quadrics (cf. [5]). The images of one point ψ in P_{2^r-1} under F are 2^r points and 2^r hyperplanes, between which strong incidence relations hold. Every hyperplane of the set contains (at least) $(2^r-1)2^{r-1}$ points of the set, and every point is contained in (at least) equally many hyperplanes of the set (because F contains that same number of null systems).

If ψ is specially chosen in one of the quadratic hypersurfaces, then the tangent hyperplane at ψ to this hypersurface belongs to the set of points and hyperplanes, and it contains in this case at least $(2^r-1)2^{r-1}+1$ points. F operates transitively on the set of points and hyperplanes, and F preserves incidence relations, hence the same (the dual) is then true of all hyperplanes (points) of the set.

In P_3 we obtain in this way (choice of ψ not special) the configuration of Kummer [9, 11, 13], consisting of 16 points and 16 planes. Each plane contains 6 points and each point is contained in 6 planes. This situation is obvious from our considerations, since if $\psi \varepsilon P_3$, then $\psi, \gamma_1\psi, \cdots, \gamma_5\psi$ lie in the plane $\Xi\psi$.

A special configuration is obtained if ψ lies on a line which intersects the invariant axes ($=$ lines) of (say) γ_1 and γ_2. Then the points $\psi, \gamma_1\psi, \gamma_2\psi, \gamma_1\gamma_2\psi$ lie on one line, and all 16 points lie on a set of four lines.

In P_7 we obtain the configuration of Barrau. If $\gamma_1, \cdots, \gamma_7$ are constructed as in (9) of Section **1**, then the points will have the same coordinates as those obtained from the device of Barrau. The configuration $Cf(64_{28})$ consists of 64 points and 64 hyperplanes in P_7, each of which is incident with (at least) 28 of the other kind. The 36 polarities in F determine 36 Q_6's in P_7. The Q_6 determined by $L(7, 8)$ is a representative of these.

Suppose now that we choose the point ψ in this Q_6. The involutions $\gamma_1, \cdots, \gamma_7$, and all other involutions of F, transform ψ into the Q_6; so that the points $\psi, \gamma_1\psi, \cdots, \gamma_7\psi$ are contained in the axis[1] $\Xi\psi$ in Q_6 (Section **3**).

Because F operates transitively on the set of points and hyperplanes, the following theorem holds:

THEOREM 11. *If* ψ *is contained in the invariant* Q_6 *of one of the* 36 *polarities, elements of the group* F *of correlations and involutions generated by* 7 *anti-commuting involutions and the related invariant polarity in* P_7, *then the images of* ψ *under* F *form a configuration* $Cf(64, 29, 8)$.

This configuration consists of 64 points, 64 hyperplanes and 64 solids (P_3's). Every hyperplane contains (at least) 29 points and (at least) 8 solids. Every point is contained in (at least) 29 hyperplanes and 8 solids. Every solid contains (at least) 8 points and is contained in (at least) 8 hyperplanes.

Room obtained this configuration in a different way [9]. Theorem 11 could also be stated as follows: The configuration of Barrau, which is special in the sense indicated above, has all incidence properties known of the configuration of Room.

In the generalization to higher dimensions, the operation Ξ determined by a linear family of involutions $L(2r+1, 2^r)$ distinguishes between different kinds of points ψ, and this distinction carries with itself a distinction between different kinds of more or less specialized generalized Kummer configurations.

THE INSTITUTE FOR ADVANCED STUDY,
PRINCETON, N. J.

REFERENCES.

[1] F. Klein, "Zur Theorie der Linienkomplexe des ersten und zweiten Grades," *Gesammelte Abhandlungen*, vol. 1 (1921), p. 53.

[2] W. Wirtinger, "Ueber das Analogen der Kummer'schen Fläche für $p = 3$," *Göttinger Nachrichten* 1889, pp. 474-480.

[3] C. M. Jessop, *A treatise on the line complex*, Cambridge, 1903, p. 23.

[4] R. W. H. T. Hudson, *Kummer's quartic surfaces*, Cambridge, 1905, p. 5.

[5] J. A. Barrau, "De uitbreiding der configuratie van Kummer op ruimten van $(2^p - 1)$afmetingen," *Koninklijke Akademie van Wetenschappen te Amsterdam*, vol. 16 (1907), pp. 205-212; "Det analogen der configuratie van Kummer in de ruimte van zeven afmetingen," *ibid.*, pp. 457-463.

[6] E. Study, "Grundlagen und Ziele der analytischen Kinematik," *Sitzungsberichte der Berliner Mathematischen Gesellschaft*, 1912, pp. 36-60.

[7] E. Cartan, "Le principe de dualité et la théorie des groupes simples et semi-simples," *Bulletin des Sciences Mathématiques*, vol. 49 (1925), pp. 361-374;

"*Leçons sur la théorie des spineurs* I, II, Paris (1938); *Géométric projective complexe*, Paris (1931).

[8] B. L. van der Waerden, "Spinoranalyse," *Göttinger Nachrichten* 1929, pp 100-109.

[9] T. G. Room, "A generalization of the Kummer 16_6 configuration (I)," *Proceedings of the London Mathematical Society*, vol. 37 (1934), pp. 292-337; "A generalization of the Kummer 16_6 configuration (II)," *ibid.*, vol. 42 (1937), pp. 529-545.

[10] O. Veblen and W. Givens, *Lectures on the geometry of complex domains*, Princeton (1936).

[11] W. Givens, "Tensor coordinates of linear spaces," *Annals of Mathematics*, vol. 38 (1937), pp. 355-385.

[12] W. van der Woude, "Ueber die Drehungsgruppe in R_6," *Annali della R. Scuola Normale Superiore di Pisa*, ser. II, vol. 4 (1935), pp. 163-175.

[13] R. Brauer and H. Weyl, "Spinors in n dimensions," *American Journal of Mathematics*, vol. 57 (1935), pp. 425-449.

[14] J. A. Schouten and J. Haantjes, "Konforme Feldtheorie II; R_6 und Spinraum," *Annali della R. Scuola Normale Superiore di Pisa*, ser. II, vol. 4 (1935), pp. 175-191.

[15] B. Segre, "Sulla totalità degli spazi di dimensione massima giacenti su di una quadrica di S_n, con particolare riguardo al caso di $n = 7$," *Bolletino del Unione Matematica Italiana*, vol. 17 (1938), pp. 1-14.

[16] E. A. Weiss, *Punktreihen Geometrie*, Leipzig, Berlin (1939).

[17] B. L. van der Waerden, *Einführung in die algebraische Geometrie*, Berlin (1939).

[18] W. L. Chow, "On the geometry of algebraic homogeneous spaces," *Annals of Mathematics*, vol. 50 (1949), pp. 32-67.

[19] J. A. Schouten, "On the geometry of spin spaces, I, II, III," *Proceedings of the Koninklijke Nederlandsche Akademie van Wetenschappen*, vol. 52 (1949), pp. 597-609; 687-695; 938-948; (to be continued).

ON THE EXISTENCE OF LAPLACE SOLUTIONS FOR LINEAR DIFFERENTIAL EQUATIONS OF SECOND ORDER.*

By AUREL WINTNER.

---

1. Let $\alpha(t)$, where

$$1 \leqq t < \infty, \tag{1}$$

be a (possibly complex-valued) function of finite total variation

$$[\alpha] = \int_1^\infty |\, d\alpha(t) \,|, \tag{2}$$

and let $f(s)$, where

$$0 \leqq s < \infty, \tag{3}$$

denote the function

$$f(s) = \int_1^\infty e^{-st} d\alpha(t), \text{ where } [\alpha] < \infty. \tag{4}$$

The latter will be chosen to be the coefficient function of the differential equation

$$x'' + f(s)x = 0, \qquad ('= d/ds), \tag{5}$$

about which the following theorem will be proved:

(i) *If* $f(s)$ *is representable, on the half-line* (3), *in the form* (4), *then* (5) *has a solution* $x(s) \not\equiv 0$ *representable, on the same half-line, in the form*

$$x(s) = \text{const.} + \int_1^\infty e^{-st} d\beta(t), \text{ where } [\beta] < \infty, \tag{6}$$

with the const. *as an arbitrary "initial value"* $x(\infty)$,

$$x(s) \to \text{const. as } s \to \infty, \tag{7}$$

where const. $= 0$ *only when* $x(s) \equiv 0$.

In some respect, (i) is related to a theorem proved in [2] for systems of first order, say

$${}^i x' = \sum_{k=1}^n f_{ik}(s)\, {}^k x, \qquad (i = 1, \cdots, n), \tag{8}$$

---

* Received July 3, 1949.

where every $f_{ik}(s)$ is of the form (5). But (5) presents an entirely different problem. In fact, if (5) is written in the form (8), as

$$\text{(8 bis)} \qquad x' = 0 \cdot x + 1 \cdot y, \qquad y' = -f(t)x + 0 \cdot y$$

$({}^1x = x,\ {}^2x = y;\ n = 2)$, then the condition imposed on the n^2 coefficient function $f_{ik}(s)$ of (8) is not satisfied by $f_{12}(s) \equiv 1$, since the latter function results by choosing $\alpha(t) = \operatorname{sgn} t$ but replacing (1) by $0 \leqq t < \infty$ in (4). Trivial examples show, however, that the results of [2] cannot apply if the range of integration reaches down to $t = 0$.

Actually, even the assertion of the theorem on (8) becomes quite different. In fact, the result of [2] is that the components ${}^ix(s)$ of every solution of (8) are of the form (6), whereas (i*) below will show that such is never the case for (5). Thus the assertion of (i) is closer to the result of [1] than to that of [2]. In fact, [1] deals with (5), but it assumes that $d\alpha(t) \leqq 0$ in (4), whilst it allows that (1) in (4) be replaced by $0 \leqq t < \infty$.

2. The fact referred to above is as follows:

(i*) *Under the assumptions of* (i), *the solution supplied by* (i) *is unique to a constant factor.*

In other words, (5) cannot have two, linearly independent, solutions of the form (6). This can be seen as follows: It is clear from (6) that $x(s) = O(1)$ and $x'(s) = o(1)$, as $s \to \infty$. Hence, if $x = u(s)$ and $x = v(s)$ are two solutions of the form (6), then their Wronskian, $u(s)v'(s) - v(s)u'(s)$, is $o(1)$. But the Wronskian of two linearly independent solutions of (5) is a non-vanishing constant and cannot, therefore, be $o(1)$. This proves (i*).

If $f(s)$ is of the form (4), let $S(\alpha)$, the "spectrum" of α, be defined as the set of those t-values which are points of non-constancy of the function $\alpha(t)$. Let $S^*(\alpha)$ denote the closure of the set of those t-values which are representable in the form $n_1t_1 + \cdots + n_mt_m$, where m and $n_1, \cdots, n_m$ denote positive integers and $t_1, \cdots, t_m$ are points of $S(\alpha)$. Then the assertion of (i) can be completed as follows:

(i bis) *Under the assumptions of* (i), (i*), *the spectrum* $S(\beta)$ *defined by* (6) *is contained in the set* $S^*(\alpha)$.

This will be clear from the proof of (i) below, if recourse is had to the "addition rule of spectra" in case of convolutions.

3. The proof of (i) depends on a process of successive approximations. Formally, the successive approximations, $x_0(s), \cdots, x_n(s), \cdots$, to the solution in question can be introduced as follows:

$$x_{n+1}(s) = 1 + \int_s^\infty (s-t)f(t)x_n(t)\,dt, \tag{9}$$

where $0 \leqq s < \infty$ and

$$x_0(s) \equiv 0. \tag{10}$$

If

$$x_{n+1}(s) = 1 + \sum_{m=1}^{n} y_m(s), \tag{11}$$

then (9) and (10) are equivalent to

$$y_{n+1}(s) = \int_s^\infty (s-t)f(t)y_n(t)\,dt, \tag{12}$$

where

$$y_0(s) \equiv 1. \tag{13}$$

It will first be shown that this recursion rule leads to a representation

$$y_n(s) = \int_1^\infty e^{-st}d\gamma_n(t), \qquad (n > 0), \tag{14}$$

where $\gamma_n(t)$ is a function satisfying $[\gamma_n] < \infty$, if the bracket is defined by (2).

According to (13) and the case $n = 0$ of (12),

$$y_1(s) = \int_s^\infty (s-t)f(t)\,dt. \tag{15}$$

If f is substituted from (4) into (15), and if the order of resulting integrations is interchanged (which, in view of $[\alpha] < \infty$, is legitimate, by absolute convergence), then it is seen from the identity

$$\int_s^\infty (s-t)e^{-tu}dt = -e^{-us}/u^2, \text{ where } 0 < u < \infty, \tag{16}$$

that (15) can be written as $y_1(s) = -\int_1^\infty e^{-st}t^{-2}d\alpha(t)$. Hence, (14) holds for $n = 1$, with

$$d\gamma_1(t) = -t^{-2}d\alpha(t). \tag{17}$$

In order to carry out the induction from n to $n+1$, suppose that (14) holds for a fixed n and for some $\gamma_n(t)$ satisfying $[\gamma_n] < \infty$. Then, by (4),

$$f(s)y_n(s) = \int_1^\infty e^{-st}d(\alpha * \gamma_n)(t), \tag{18}$$

where the asterisk denotes the operation of convolution, that is,

$$(\lambda * \mu)(t) = \int_1^t \lambda(t-u)\,d\mu(u) = (\mu * \lambda)(t),$$

if $[\lambda] < \infty$ and $[\mu] < \infty$. But if fy_n is substituted from (18) into (12), and if (16) is then used in the same way as above, it is seen that

$$y_{n+1}(s) = -\int_1^\infty e^{-st}t^{-2}d(\alpha * \gamma_n)(t).$$

Accordingly, n in (14) can be replaced by $n+1$ if $\gamma_{n+1}(t)$ is defined, corresponding to (17), by

$$d\gamma_{n+1}(t) = -t^{-2}d(\alpha * \gamma_n)(t). \tag{19}$$

This completes the induction.

4. Needless to say, γ_n can be left undetermined at its points of discontinuity and to an additive constant, since $[\gamma_n] < \infty$. For instance, it can be assumed that

$$\gamma_n(1) = 0, \text{ and } \gamma_n(t-0) = \gamma_n(t) \text{ if } 1 < t < \infty. \tag{20}$$

It turns out that $\gamma_n(t)$ is independent of t if $1 \leqq t < n$ or, in view of the first of the relations (20), that

$$\gamma_n(t) = 0 \text{ if } 1 \leqq t < n, \text{ where } n > 1;\ \gamma_1(1) = 0. \tag{21}$$

This follows from (17) and (19), where $1 \leqq t < \infty$, by an induction from n to $n+1$, if use is made of the "addition rule of spectra." In fact, this induction is exactly the same as that given in [2], pp. 335-336, and will therefore be omitted.

On the other hand, it is readily seen from the definitions of the symbols

$[\ ]$, $*$ that $[\lambda * \mu] \leqq [\lambda][\mu]$. It follows therefore from (19) and (21) that $[\gamma_{n+1}] \leqq (n+1)^{-2}[\alpha][\gamma_n]$. This recursive inequality, when applied to $n = 1, 2, \cdots$, clearly implies that $[\gamma_n] = O(a^n/n!^2)$ as $n \rightarrow \infty$, where a is independent of n. In particular

$$\sum_{n=1}^{\infty} [\gamma_n] < \infty. \tag{22}$$

It is clear from (22) and from the first of the relations (20) that the series

$$\beta(t) = \sum_{n=1}^{\infty} \gamma_n(t), \text{ where } 1 \leqq t < \infty, \tag{23}$$

defines a function satisfying $\beta(1) = 0$ and $[\beta] < \infty$. It is also seen that the function (23) is such as to satisfy, for $0 \leqq s < \infty$, the limit relation

$$\sum_{m=1}^{n} \int_1^{\infty} e^{-st} d\gamma_m(t) \rightarrow \int_1^{\infty} e^{-st} d\beta(t), \text{ as } n \rightarrow \infty.$$

In view of (14) and (11), this means that

$$x_n(s) \rightarrow x(s) \text{ as } n \rightarrow \infty;\ 0 \leqq s < \infty, \tag{24}$$

if $x(s)$ is defined by

$$x(s) = 1 + \int_1^{\infty} e^{-st} d\beta(t) \qquad ([\beta] < \infty). \tag{25}$$

Finally, it is readily seen from (22), (23) and (4) that the limit process (24) can be carried out beneath the integral sign on the right of (9). This means that the function (25) is a solution of the integral equation

$$x(s) = 1 + \int_s^{\infty} (s-t) f(t) x(t)\, dt; \qquad 0 \leqq s < \infty. \tag{26}$$

Two differentiations of (26) show that the function $x(s)$ is a solution of (5). In addition, it is clear from (25) that $x(s)$ satisfies the case const. $= 1$ of (7), (6). Since a solution of (5) can be multiplied by an arbitrary constant, it follows that, in order to complete the proof of (i), only the last assertion of (i), that following (7), remains to be ascertained. But the truth of this assertion is contained in the uniqueness statement of (i*), which was verified above.

5. Nothing is changed in the above proofs if the half-line (3) is replaced by the half-plane $\sigma \geqq 0$, where $s = \sigma + it$, $-\infty < t < \infty$. In particular, if the claims of (i), (i*), (i bis) are applied on the boundary of the half-plane $\sigma \geqq 0$, there results the following theorem:

(ii) *Let a continuous* $f(t)$, *where* $-\infty < t < \infty$, *be a function representable in the form*

$$f(t) = \int_1^\infty e^{its} d\alpha(s), \text{ where } \int_1^\infty |\, d\alpha(s) \,| < \infty. \tag{27}$$

Then, corresponding to every constant c, *the differential equation*

$$x'' + f(t)x = 0 \qquad (' = d/dt) \tag{28}$$

has, for $-\infty < t < \infty$, *a unique solution of the form*

$$x(t) = c + \int_1^\infty e^{its} d\beta(s), \text{ where } \int_1^\infty |\, d\beta(s) \,| < \infty. \tag{29}$$

The spectrum $S(\beta)$ *is contained in the set* $S^*(\alpha)$.

Obviously, the lower limit, 1, of the integrations (27), (29) can be replaced by any positive number.

Of particular interest is the case in which the function α is a step-function. Then (ii) leads to the following theorem:

(iii) *Let* $f(t)$ *be a uniformly almost-periodic function of the form*

$$f(t) \sim \sum_{n=1}^\infty a_n e^{i\lambda_n t}, \text{ where } \lambda_n > \text{const.} > 0, \tag{30}$$

and suppose that

$$\sum_{n=1}^\infty |\, a_n \,| < \infty. \tag{31}$$

Then the differential equation (28) *has a uniformly almost-periodic solution of the form*

$$x(t) \sim c + \sum_{n=1}^\infty b_n e^{i\mu_n t}, \qquad (\mu_n > \text{const.} > 0), \tag{32}$$

where every exponent μ_n *is a linear combination, with positive integral coefficients, of the exponents* λ_n, *and*

$$\sum_{n=1}^\infty |\, b_n \,| < \infty. \tag{33}$$

Furthermore, the assignment of an arbitrary mean-value

$$c = \lim_{v-u\to\infty} (v-u)^{-1} \int_u^v x(t)\,dt$$

as an integration constant determines the solution (32) *uniquely.*

6. In view of the uniqueness theorem expressed by (i*), and of the corresponding assertion of (ii), there arises the question as to the form of another (linearly independent) solution or, equivalently, of the general solution of the differential equations (5), (28) under the respective assumptions of (i), (ii). It will be sufficient to develop the answer to this question for the case of (i), since it then follows by the above device, $s = it$, for the cases of (ii) and (iii). In the case of (i), the theorem is as follows:

(I) *If* $f(s)$ *is representable, on the half-line* (3), *in the form* (4), *then the general solution of* (5) *is* $x(s) = c_1x_1(s) + c_2x_2(s)$, *where* c_1, c_2 *are arbitrary constants, and* $x_1(s)$, $x_2(s)$ *solutions representable, on the half-line* (3), *as*

$$(34) \qquad x_1(s) = 1 + \int_1^\infty e^{-st}d\beta(t)\,;\ [\beta] < \infty, \text{ and } x_2(s) = sx_1(s) + y(s),$$

respectively, where $y(s)$ *is a function of the form*

$$(35) \qquad y(s) = \int_1^\infty e^{-st}d\gamma(t)\,; \qquad [\gamma] < \infty.$$

First, $x_1(s)$ in (34) is the solution (25), supplied by (i). On the other hand, substitution of $x_2(s)$ from (34) into (5) gives

$$(36) \qquad y'' + f(s)y = g(s),$$

where

$$(37) \qquad g(s) = -2x_1'(s),$$

since

$$(38) \qquad x_1''(s) = -f(s)x_1(s),$$

by (5). Hence, in order to prove (I), it is sufficient to show that the inhomogeneous differential equation (36) has, in the case (37), a solution of the form (35). In fact, since the definite integrals occurring in (34) and (35) tend to 0 as $s \to \infty$, it is clear from (34) that $x_1(s) \to 1$ but $|x_2(s)| \to \infty$, which assures that $x_2(s)$ is linearly independent of $x_1(s)$.

In order to prove for (36) the existence of a solution (35), use will be made of the fact that the function (37) is representable in the form

$$(39) \qquad g(s) = \int_1^\infty e^{-st} d\phi(t); \qquad [\phi] < \infty.$$

This can be verified as follows:

By (4) and the first of the relations (35),

$$f(s)x_1(s) = \int_1^\infty e^{-st} d\alpha(t) + \int_1^\infty e^{-st} d\alpha * \beta(t).$$

On the other hand, from (38) and the first of the relations (35),

$$\int_1^\infty e^{-st} t^2 d\beta(t) = -f(s)x_1(s).$$

Consequently, $t^2 d\beta(t) = -d\alpha(t) - d\alpha*\beta(t)$. In view of $[\alpha*\beta] \leqq [\alpha][\beta]$, where $[\alpha] < \infty$ and $[\beta] < \infty$, this implies, by (1), that

$$\int_1^\infty t^2 \, |d\beta(t)| < \infty; \text{ in particular, } \int_1^\infty |t d\beta(t)| < \infty.$$

It follow therefore from (37) and from the first of the relations (34) that (39) is satisfied by $d\phi(t) = -2t d\beta(t)$.

Accordingly, more than what is needed for (I) is contained in the following theorem:

(II) *If* $f(s)$, $g(s)$ *are functions representable, on the half-line* (3), *as integrals of the form* (4), (39), *then the inhomogeneous differential equation* (36) *has, on the same half-line, a solution of the form* (35).

The proof of (II) proceeds again by successive approximations, except that (9) must be replaced by

$$y_{n+1}(s) = h(s) + \int_1^\infty (s-t) f(t) y_n(t) \, dt, \text{ where } h(s) = \int_1^\infty e^{-st} t^{-2} d\gamma(t),$$

and (10) by $y_0(s) \equiv 0$ or $y_1(s) = h(s)$. Due to the assumption (34), the induction from n to $n+1$ and the limit process $n \rightarrow \infty$ can then be carried out in exactly the same way as in the proof of (i), and will not therefore be repeated.

Clearly, both (I) and (II) can be replaced by the theorems which relate to (I) and (II) in the same way as either (ii) or (iii) related to (i).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] A. Wintner, "On the Laplace-Fourier transcendents occurring in mathematical physics," *American Journal of Mathematics*, vol. 69 (1947), pp. 87-98.

[2] ———, "On the classical existence theorem of ordinary differential equations," *ibid.*, vol. 71 (1949), pp. 331-338.

ISOMORPHIC GROUPS OF LINEAR TRANSFORMATIONS.*[1]

By C. E. RICKART.

Introduction. Consider a system $\{\mathfrak{X}, \mathfrak{D}, \mathfrak{X}^* : \mathfrak{G}\}$ of the following type: $\mathfrak{X}$, $\mathfrak{X}^*$ are dual [3, p. 15] (right and left respectively) linear vector spaces over the division ring $\mathfrak{D}$ (not of characteristic 2) with the dimension of $\mathfrak{X}$ over $\mathfrak{D}$ greater than two; $\mathfrak{G}$ is a family of linear transformations on $\mathfrak{X}$ to $\mathfrak{X}$ each of which possesses an adjoint on $\mathfrak{X}^*$ [3, p. 16]; $\mathfrak{G}$ is a group under the "circle operation" $A \circ B = A + B - AB$; $\mathfrak{G}$ contains all one-dimensional involutions which possess adjoints. Denote by $\{\mathfrak{Y}, \mathfrak{E}, \mathfrak{Y}^* : \mathfrak{H}\}$ a second such system and assume that $\mathfrak{G}$ and $\mathfrak{H}$ are isomorphic as groups. In this situation, two principal results are obtained. The first states that one of the following pairs of linear space isomorphisms must hold: (i) $\mathfrak{X} \sim \mathfrak{Y}$ and $\mathfrak{X}^* \sim \mathfrak{Y}^*$ or (ii) $\mathfrak{X} \sim \mathfrak{Y}^*$ and $\mathfrak{X}^* \sim \mathfrak{Y}$. The second states that the group isomorphism is essentially generated by the corresponding linear space isomorphism.

G. Mackey [5, p. 251] has obtained the first result for the special case in which $\mathfrak{X}$, $\mathfrak{Y}$ are infinite dimensional normed linear spaces over the real numbers; $\mathfrak{X}^*$, $\mathfrak{Y}^*$ are the conjugate spaces of $\mathfrak{X}$, $\mathfrak{Y}$ respectively and $\mathfrak{G}$, $\mathfrak{H}$ are the (multiplicative) groups of all linear transformations on $\mathfrak{X}$, $\mathfrak{Y}$ respectively which are continuous and possess continuous inverses. Our proof is based on an extension of methods used by Mackey. The second result reduces essentially to a result of Dieudonné [1, Théorème 1] in the finite dimensional case. The finite dimensional case over a field has also been treated by Schreier and van der Waerden [6, p. 305].

In Section **1** are collected a few observations on dual spaces and involutions. Involutions are studied in detail in Sections **2**, **3**. The first result mentioned above is proved in Section **4** and the second is proved in Section **5**.

The groups studied here reduce essentially to full linear groups in the finite dimensional case. In a subsequent paper we will carry out a similar investigation for groups which reduce in finite dimensions to unitary, orthogonal or symplectic groups.

We take this opportunity to acknowledge our indebtedness to N. Jacobson for a number of suggestions which led to substantial improvement of our

---

* Received August 26, 1949.

[1] Prepared in part under the sponsorship of the Office of Naval Research. Presented to the American Mathematical Society, September, 1948 and September, 1949.

results. In particular, a suggestion of his enabled us to complete the proof of Theorem 5. 1.

1. Dual linear vector spaces. Involutions. Let $\mathfrak{X}$ denote a linear vector space over a division ring $\mathfrak{D}$. The space $\mathfrak{X}$ is called a *right* (*left*) $\mathfrak{D}$-*space* if the scalar multiplication is written on the right (left). A right $\mathfrak{D}$-space $\mathfrak{X}$ is said to be *isomorphic* to a right (left) $\mathcal{E}$-space $\mathfrak{Y}$ provided there exists a one-to-one addition-preserving correspondence $x \leftrightarrow y$ between $\mathfrak{X}$ and $\mathfrak{Y}$ and an isomorphism (anti-isomorphism) $\alpha \leftrightarrow \beta$ between $\mathfrak{D}$ and $\mathcal{E}$ such that $x\alpha \leftrightarrow y\beta$ ($x\alpha \leftrightarrow \beta y$). A similar definition holds with "right" and "left" interchanged.

Let $\mathfrak{X}$ and $\mathfrak{X}^*$ be right and left $\mathfrak{D}$-spaces respectively. Then $\mathfrak{X}$ and $\mathfrak{X}^*$ are said to be dual [3, p. 15] provided there exists a function (x^*, x) on $\mathfrak{X}^* \times \mathfrak{X}$ to $\mathfrak{D}$ with the following properties: (1) $(x^*, x) = 0$ for every $x^* \varepsilon \mathfrak{X}^*$ ($x \varepsilon \mathfrak{X}$) implies $x = 0$ ($x^* = 0$), (2) $(x^*, x\lambda + y\mu) = (x^*, x)\lambda + (x^*, y)\mu$ and $(\lambda x^* + \mu y^*, x) = \lambda(x^*, x) + \mu(y^*, x)$ for arbitrary $\lambda, \mu \varepsilon \mathfrak{D}$, $x, y \varepsilon \mathfrak{X}$ and $x^*, y^* \varepsilon \mathfrak{X}^*$.

Consider a subset $S \subset \mathfrak{X}$ ($S \subset \mathfrak{X}^*$) and define $S^\perp$ as the set of all $x^* \varepsilon \mathfrak{X}^*$ ($x \varepsilon \mathfrak{X}$) such that $(x^*, x) = 0$ for every $x \varepsilon S$ ($x^* \varepsilon S$). If the set S is finite, then the set $x^* + S^\perp$ ($x + S^\perp$) is called a neighborhood of $x^* \varepsilon \mathfrak{X}^*$ ($x \varepsilon \mathfrak{X}$). Under these neighborhood systems, $\mathfrak{X}^*$ and $\mathfrak{X}$ become T_1-spaces [3, p. 16]. If S consists of only one non-zero element, then $S^\perp$ is a maximal linear subspace of $\mathfrak{X}^*$ (or $\mathfrak{X}$). Let $x_1, \cdots, x_n \varepsilon \mathfrak{X}$ ($x^*_1, \cdots, x^*_n \varepsilon \mathfrak{X}^*$) be arbitrary linearly independent elements; then there exist elements $x^*_1, \cdots, x^*_n \varepsilon \mathfrak{X}^*$ ($x_1, \cdots, x_n \varepsilon \mathfrak{X}$) such that $(x^*_i, x_j) = \delta_{ij}$ [3, p. 16].

Let A be a linear transformation on $\mathfrak{X}$ to $\mathfrak{X}$. A linear transformation A^* on $\mathfrak{X}^*$ to $\mathfrak{X}^*$ is said to be *adjoint* to A provided $(x^*, xA) = (x^*A^*, x)$ for all $x \varepsilon \mathfrak{X}$ and $x^* \varepsilon \mathfrak{X}^*$. Whenever A^* exists it is unique, and a necessary and sufficient condition for the existence of A^* is that A be continuous in the T_1-topology defined above [3, p. 17]. The linear transformation A is said to be finite dimensional provided its range $\mathfrak{X}A$ is finite dimensional. Every continuous finite dimensional linear transformation A has the form $xA = \sum_{i=1}^{n} a_i(a^*_i, x)$, where $a_i \varepsilon \mathfrak{X}$ and $a^*_i \varepsilon \mathfrak{X}^*$ are fixed and n is the dimension of $\mathfrak{X}A$. Conversely, every finite dimensional transformation of this form is continuous and $x^*A^* = \sum_{i=1}^{n} (x^*, a_i)a^*_i$ defines its adjoint [3, p. 17].

In addition to the usual operations of addition and multiplication for linear transformations on a linear space, we also consider the *circle operation* $A \circ B = A + B - AB$ [4, p. 153; 2, p. 455]. This operation is associative and has the identically zero transformation as an identity. The family of all continuous linear transformations on $\mathfrak{X}$ which have continuous inverses relative to the circle operation constitute a group under this operation. Moreover, under the correspondence $A \leftrightarrow I - A$, this group is isomorphic to the multiplicative group of all continuous linear transformations which have continuous inverses relative to ordinary multiplication. The inverse of a linear transformation A relative to the circle operation will be denoted by A°.

A linear transformation T is called an *involution* (relative to the circle operation) provided $T \circ T = 0$ or, equivalently, $T^2 = T2$. Observe that the adjoint of an involution is also an involution. For what follows, it is necessary to require the characteristic of $\mathfrak{D}$ to be different from 2.

1. 1 THEOREM. *For each involution T on $\mathfrak{X}$ there exists a unique decomposition, $\mathfrak{X} = \mathfrak{M} + \mathfrak{N}$, $\mathfrak{M} \cap \mathfrak{N} = (0)$, of $\mathfrak{X}$ such that $xT = x2$ for $x \varepsilon \mathfrak{M}$ and $xT = 0$ for $x \varepsilon \mathfrak{N}$. If T^* exists, then the analogous decomposition of $\mathfrak{X}^*$ associated with T^* is $\mathfrak{X}^* = \mathfrak{N}^{\perp} + \mathfrak{M}^{\perp}$.*

Define $E = T\frac{1}{2}$ and observe that $E^2 = E$; that is, E is a projection on $\mathfrak{X}$. It is evident that the desired decomposition of $\mathfrak{X}$ is given by $\mathfrak{M} = \mathfrak{X}E$ and $\mathfrak{N} = \mathfrak{X}(I - E)$. If T^* exists, then so also does E^*; in fact $E^* = \frac{1}{2}T^*$. Since $x^*E^* = 0$ if, and only if, $(x^*, xE) = 0$ for every $x \varepsilon \mathfrak{X}$, it follows that $x^*E^* = 0$ is equivalent to $x^* \varepsilon \mathfrak{M}^{\perp}$. Also, since $x^*E^* = x^*$ if, and only if, $(x^*, x - xE) = 0$ for every $x \varepsilon \mathfrak{X}$, it follows that $x^*E^* = x^*$ is equivalent to $x^* \varepsilon \mathfrak{N}^{\perp}$. This completes the proof.

The subspaces $\mathfrak{M}$, $\mathfrak{N}$ associated with T by Theorem 1. 1 are called the *subspaces* of T. The involution is said to be *minimal* provided $\mathfrak{M}$ is one-dimensional and *maximal* provided $\mathfrak{N}$ is one-dimensional. In either case, it is said to be *extremal*. Observe that the operation $T \circ I2$ interchanges the subspaces of T. Thus, if T is minimal (maximal), then $T \circ I2$ is maximal (minimal). If T_1, T_2 are two involutions with the same subspaces, then either $T_2 = T_1$ or $T_2 = T_1 \circ I2$.

The following lemma is an easy consequence of the definition of involution and the representation of continuous finite dimensional transformations mentioned above.

1. 2. LEMMA. *Let T be a continuous involution on $\mathfrak{X}$. Then T is minimal if, and only if, it is of the form $xT = t2(t^*, x)$ and is maximal if, and only if, it is of the form $xT = x2 - t2(t^*, x)$, where $t \varepsilon \mathfrak{X}$ and $t^* \varepsilon \mathfrak{X}^*$ are fixed such that $(t^*, t) = 1$.*

Throughout the remainder of this paper, $\mathfrak{X}$ and $\mathfrak{X}^*$ will be dual $\mathfrak{D}$-spaces where the characteristic of $\mathfrak{D}$ is different from 2 and $\dim(\mathfrak{X}) \geq 3$. Also $\mathscr{G}$ will be a family of continuous linear transformations on $\mathfrak{X}$ which is a group under the circle operation and which contains all continuous minimal involutions. The family $\mathscr{G}^*$ of adjoints of elements of $\mathscr{G}$ is also a group under the circle operation. Moreover, the groups $\mathscr{G}$ and $\mathscr{G}^*$ are anti-isomorphic under the natural correspondence $A \leftrightarrow A^*$. A system of the type described here will be denoted by $\{\mathfrak{X}, \mathfrak{D}, \mathfrak{X}^* : \mathscr{G}\}$.

2. Intersection properties of extremal involutions. Consider the system $\{\mathfrak{X}, \mathcal{E}, \mathfrak{X}^* : \mathscr{G}\}$ and denote by $\mathscr{I}$ the set of all involutions in $\mathscr{G}$. The objective of this section is to obtain, in terms only of the group operation in $\mathscr{G}$, a necessary and sufficient condition that two non-commutative extremal involutions in $\mathscr{I}$ shall have a subspace in common.

Let $\mathscr{S}$ be an arbitrary subset of $\mathscr{I}$ and denote by $c(\mathscr{S})$ the set of all $T \varepsilon \mathscr{I}$ such that $T \circ S = S \circ T$ for every $S \varepsilon \mathscr{S}$. Observe that $c(\mathscr{S})$ can also be defined as the set of all $T \varepsilon \mathscr{I}$ such that $TS = ST$ for every $S \varepsilon \mathscr{S}$, since $TS = ST$ is equivalent to $T \circ S = S \circ T$.

2.1. Lemma. *Let T be an involution with subspaces $\mathfrak{M}$, $\mathfrak{N}$ and let A be an arbitrary linear transformation on $\mathfrak{X}$. Then $T \circ A = A \circ T$ if, and only if, $\mathfrak{M}A \subseteq \mathfrak{M}$ and $\mathfrak{N}A \subseteq \mathfrak{N}$.*

The necessity of the condition is trivial. Therefore assume $\mathfrak{M}A \subseteq \mathfrak{M}$, $\mathfrak{N}A \subseteq \mathfrak{N}$ and, for arbitrary $x \varepsilon \mathfrak{X}$, write $x = m + n$ where $m \varepsilon \mathfrak{M}$ and $n \varepsilon \mathfrak{N}$. Then $xTA = mA2$. Also, since $mA \varepsilon \mathfrak{M}$ and $nA \varepsilon \mathfrak{N}$, $xAT = mAT + nAT = mA2$. Therefore $TA = AT$ and hence $T \circ A = A \circ T$.

2.2 Lemma. *Let T_i be involutions with subspaces $\mathfrak{M}_i$, $\mathfrak{N}_i$ $(i = 1, 2)$. A sufficient condition for $T_1 \circ T_2 = T_2 \circ T_1$ is that either $\mathfrak{M}_1 \subseteq \mathfrak{M}_2$ and $\mathfrak{N}_2 \subseteq \mathfrak{N}_1$ or $\mathfrak{M}_1 \subseteq \mathfrak{N}_2$ and $\mathfrak{M}_2 \subseteq \mathfrak{N}_1$. If T_1 is minimal, then the condition is also necessary.*

Assume first that $\mathfrak{M}_1 \subseteq \mathfrak{M}_2$ and $\mathfrak{N}_2 \subseteq \mathfrak{N}_1$. Obviously $\mathfrak{M}_1 T_2 \subseteq \mathfrak{M}_1$. Moreover, let $n_1 \varepsilon \mathfrak{N}_1$ and write $n_1 = m_2 + n_2$, where $m_2 \varepsilon \mathfrak{M}_2$ and $n_2 \varepsilon \mathfrak{N}_2$. Since $n_2 \varepsilon \mathfrak{N}_1$, we have $m_2 T_1 = 0$. Therefore, $n_1 T_2 T_1 = m_2 2 T_1 = 0$, so that $n_1 T_2 \varepsilon \mathfrak{N}_1$. In other words $\mathfrak{N}_1 T_2 \subseteq \mathfrak{N}_1$. That $T_1 \circ T_2 = T_2 \circ T_1$ now follows from Lemma 2.1. Sufficiency of the second condition is proved similarly. Now let $T_1 \circ T_2 = T_2 \circ T_1$ and assume T_1 minimal. By Lemma 2.1, $\mathfrak{M}_1 T_2 \subseteq \mathfrak{M}_1$. Therefore, since $\mathfrak{M}_1$ is one-dimensional, if u is a non-zero element of $\mathfrak{M}_1$, then $uT_2 = u\lambda$ where $\lambda \varepsilon \mathfrak{D}$. Since $T_2{}^2 = T_2 2$, this implies either $\lambda = 2$ and $\mathfrak{M}_1 \subseteq \mathfrak{M}_2$ or $\lambda = 0$ and $\mathfrak{M}_1 \subseteq \mathfrak{N}_2$. Suppose $\mathfrak{M}_1 \subseteq \mathfrak{M}_2$ and let n_2 be an arbitrary element of $\mathfrak{N}_2$. Write $n_2 = m_1 + n_1$, where $m_1 \varepsilon \mathfrak{M}_1$ and $n_1 \varepsilon \mathfrak{N}_1$. Then $0 = n_2 T_2 T_1 = n_2 T_1 T_2 = m_1 2 T_2 = m_1 4$. Therefore $m_1 = 0$ and hence $\mathfrak{N}_2 \subseteq \mathfrak{N}_1$. A similar proof holds in the case $\mathfrak{M}_1 \subseteq \mathfrak{N}_2$.

2.3 COROLLARY. *If T_1, T_2 are both minimal and commute, then either $T_1 = T_2$ or $T_1 T_2 = 0$.*

2.4 LEMMA. *Let T_i be non-commutative minimal involutions in $\mathfrak{I}$ with subspaces $\mathfrak{M}_i, \mathfrak{N}_i$ $(i = 1, 2)$. A necessary and sufficient condition for T to belong to $c(T_1, T_2)$ is that one subspace of T contain $\mathfrak{M}_1 + \mathfrak{M}_2$ and the other be contained in $\mathfrak{N}_1 \cap \mathfrak{N}_2$.*

The sufficiency is immediate from the first part of Lemma 2.2. On the other hand, if $T \varepsilon c(T_1, T_2)$ and has subspaces $\mathfrak{M}$, $\mathfrak{N}$, then by the second part of Lemma 2.2, either $\mathfrak{M}_i \subseteq \mathfrak{M}$ and $\mathfrak{N} \subseteq \mathfrak{N}_i$ or $\mathfrak{M}_i \subseteq \mathfrak{N}$ and $\mathfrak{M} \subseteq \mathfrak{N}_i$ for $i = 1, 2$. We have only to show that the same case occurs for both $i = 1, 2$. Therefore suppose $\mathfrak{M}_1 \subseteq \mathfrak{M}$, $\mathfrak{N} \subseteq \mathfrak{N}_1$, $\mathfrak{M}_2 \subseteq \mathfrak{N}$ and $\mathfrak{M} \subseteq \mathfrak{N}_2$. But this implies $\mathfrak{M}_1 \subseteq \mathfrak{N}_2$ and $\mathfrak{M}_2 \subseteq \mathfrak{N}_1$ so that T_1, T_2 commute, contrary to hypothesis.

2.5 LEMMA. *Let T_1, T_2 be non-commutative extremal involutions in $\mathfrak{I}$. A sufficient condition for $T \varepsilon c(c(T_1, T_2))$ is that one subspace of T be contained in the union of the one-dimensional subspaces of T_1, T_2 and the other contain the intersection of their non-one-dimensional subspaces. The condition is also necessary if* $\dim(\mathfrak{X}) > 3$ *or if* $\dim(\mathfrak{X}) = 3$ *and T_1, T_2 have a common subspace.*

Observe that there is no loss in assuming the T_i to be minimal with subspaces $\mathfrak{M}_i$, $\mathfrak{N}_i$ $(i = 1, 2)$. Let $T \varepsilon \mathfrak{I}$ satisfy the given condition and consider an arbitrary $T' \varepsilon c(T_1, T_2)$. By Lemma 2.4, one subspace of T' contains $\mathfrak{M}_1 + \mathfrak{M}_2$ and the other is contained in $\mathfrak{N}_1 \cap \mathfrak{N}_2$. That T commutes with T' and is therefore in $c(c(T_1, T_2))$ follows by Lemma 2.1. This proves the sufficiency.

Now assume $\dim(\mathfrak{X}) > 3$ and observe that, sinve $\mathfrak{N}_1$ and $\mathfrak{N}_2$ are maximal, $\dim(\mathfrak{N}_1 \cap \mathfrak{N}_2) \geqq 2$. Moreover, since $\dim(\mathfrak{M}_1 + \mathfrak{M}_2) \leqq 2$ and $\mathfrak{M}_1 + \mathfrak{M}_2$ cannot be contained in $\mathfrak{N}_1 \cap \mathfrak{N}_2$, the set $\mathfrak{P}$ of elements of $\mathfrak{N}_1 \cap \mathfrak{N}_2$ which are not in $\mathfrak{M}_1 + \mathfrak{M}_2$ is non-vacuous. Also, if $z \varepsilon \mathfrak{P}$, then $z \neq 0$ and $z + \mathfrak{N}_1 \cap \mathfrak{N}_2 \cap (\mathfrak{M}_1 + \mathfrak{M}_2) \subseteq \mathfrak{P}$. It follows immediately that the linear subspace generated by $\mathfrak{P}$ must equal $\mathfrak{N}_1 \cap \mathfrak{N}_2$. Now take $T \varepsilon c(c(T_1, T_2))$ and let $\mathfrak{M}$, $\mathfrak{N}$ be its subspaces. Also let y, z be arbitrary elements of $\mathfrak{X}$ such that $y \not\varepsilon \mathfrak{M}_1 + \mathfrak{M}_2$ and $z \varepsilon \mathfrak{P}$. Since $\mathfrak{M}_1$, $\mathfrak{M}_2$ are one-dimensional and $y, z \not\varepsilon \mathfrak{M}_1 + \mathfrak{M}_2$, there exists $z^* \varepsilon \mathfrak{X}^*$ such that $(z^*, \mathfrak{M}_1 + \mathfrak{M}_2) = (0)$, $(z^*, z) = 1$, $(z^*, y) \neq 0$. Define Z by $xZ = z2(z^*, x)$. Then, by Lemma

2. 4, $Z \varepsilon c(T_1, T_2)$ so that Z and T commute. By Lemma 2. 2, one subspace of T must contain z and (z^*, x) must vanish for x in the other subspace, which must therefore exclude y. Holding z fixed in $\mathfrak{P}$ and allowing y to vary over elements of $\mathfrak{X}$ not in $\mathfrak{M}_1 + \mathfrak{M}_2$, we conclude that one subspace of T is contained in $\mathfrak{M}_1 + \mathfrak{M}_2$. On the other hand, by varying z over $\mathfrak{P}$, we see that the other subspace of T must contain $\mathfrak{P}$ and hence $\mathfrak{N}_1 \cap \mathfrak{N}_2$, by linearity. This completes the proof of the necessity for the case $\dim(\mathfrak{X}) > 3$.

If $\dim(\mathfrak{X}) = 3$, then it can happen that the set $\mathfrak{P}$ is vacuous. In this case $c(T_1, T_2)$ contains only 0 and $I2$ so that $c(c(T_1, T_2)) = \mathfrak{A}$ and the necessity obviously fails. On the other hand, if $\mathfrak{P}$ is not vacuous, the above proof of the necessity goes through. In particular, if T_1 and T_2 have a common subspace, then $\mathfrak{P}$ cannot be vacuous. This completes the proof.

2. 6 THEOREM. *Let T_1, T_2 be arbitrary non-commutative extremal involutions in $\mathfrak{A}$. Then a necessary and sufficient condition for T_1, T_2 to have a subspace in common is that $c(c(U_1, U_2)) = c(c(T_1, T_2))$ for every pair of non-commutative extremal involutions $U_1, U_2 \varepsilon c(c(T_1, T_2))$.*

There is no loss in assuming T_i to be minimal with subspaces $\mathfrak{M}_i$, $\mathfrak{N}_i$ and written in the form $xT_i = t_i 2(t^*_i, x)$, $(t^*_i, t) = 1$ $(i = 1, 2)$. Note that T_1, T_2 will have a common subspace if, and only if, the elements in one, and only one, of the pairs t_1, t_2 and t^*_1, t^*_2 are linearly dependent.

Assume first that T_1, T_2 do not have a subspace in common. Define $u_1 = t_1 + t_2 - t_2(t^*_2, t_1)$; then u_1, t_2 are linearly independent and $(t^*_2, u_1) = 1$. Now define U_1 by $xU_1 = u_1 2(t^*_2, x)$ and take $U_2 = T_2$. If $\mathfrak{M}'_i$, $\mathfrak{N}'_i$ are the subspaces of U_i, then $\mathfrak{M}'_i \subseteq \mathfrak{M}_1 + \mathfrak{M}_2$ and $\mathfrak{N}'_i = \mathfrak{N}_2 \supseteq \mathfrak{N}_1 \cap \mathfrak{N}_2$ $(i = 1, 2)$. Hence, by Lemma 2. 5, $U_i \varepsilon c(c(T_1, T_2))$ $(i = 1, 2)$. However, since $\mathfrak{N}'_i \cap \mathfrak{N}'_2 = \mathfrak{N}_2$ and $\mathfrak{N}_1 \neq \mathfrak{N}_2$, Lemma 2.5 also gives $T_1 \not\varepsilon c(c(U_1, U_2))$. Since $T_1 \varepsilon c(c(T_1, T_2))$, this proves the sufficiency.

Now assume that T_1, T_2 have a common subspace, say $\mathfrak{N}_1 = \mathfrak{N}_2$. Then t_1, t_2 are linearly independent. Let U_1, U_2 be any pair of non-commutative minimal involutions in $c(c(T_1, T_2))$ and let $xU_i = u_i 2(u^*_i, x)$ $(i = 1, 2)$. If $\mathfrak{M}'_i$, $\mathfrak{N}'_i$ are the subspaces of U_i $(i = 1, 2)$, it follows by Lemma 2. 5 that $\mathfrak{N}'_i \supseteq \mathfrak{N}_1 \cap \mathfrak{N}_2$. Therefore $\mathfrak{N}'_1 = \mathfrak{N}'_2 = \mathfrak{N}_1 = \mathfrak{N}_2$. Also $\mathfrak{M}'_1 + \mathfrak{M}'_2 \subseteq \mathfrak{M}_1 + \mathfrak{M}_2$. But U_1, U_2 do not commute; therefore $\dim(\mathfrak{M}'_1 + \mathfrak{M}'_2) = 2$ so that $\mathfrak{M}'_1 + \mathfrak{M}'_2 = \mathfrak{M}_1 + \mathfrak{M}_2$. Another application of Lemma 2. 5 gives $c(c(U_1, U_2)) = c(c(T_1, T_2))$. On the other hand, if $\mathfrak{M}_1 = \mathfrak{M}_2$, then the above argument applied to the adjoints T^*_i, U^*_i gives $c(c(U^*_1, U^*_2)) = c(c(T^*_1, T^*_2))$ and this evidently implies $c(c(U_1, U_2)) = c(c(T_1, T_2))$, completing the proof.

3. Group characterization of extremal involutions. The objective here is to characterize an extremal involution of $\mathfrak{I}$ in terms only of the group operation in $\mathfrak{G}$. Let $T_1, T_2 \varepsilon \mathfrak{I}$ and denote by $\nu(T_1, T_2)$ the number of distinct elements in $c(c(T_1, T_2))$. Define

$$\nu = \max \nu(T_1, T_2),\ T_1 \circ T_2 = T_2 \circ T_1; \qquad \nu_T = \max \nu(T, T'),\ T \circ T' = T' \circ T.$$

3.1 Lemma. *Let P be a continuous projection on $\mathfrak{X}$ with* $\dim(\mathfrak{X}P) > 1$ *and let Z be any linear transformation on $\mathfrak{X}$ which commutes with every minimal involution $U \varepsilon \mathfrak{I}$ such that $UP = PU = U$. Then there exists ζ in the center of $\mathfrak{D}$ such that $PZ = P\zeta$. If Z is an involution, then $\zeta = 2$ or 0 and furthermore the condition on the dimension of $\mathfrak{X}P$ can be dropped.*

Let u be an arbitrary non-zero element of $\mathfrak{X}P$ and choose $v^* \varepsilon \mathfrak{X}^*$ such that $(v^*, u) = 1$. Define $u^* = v^*P^*$; then also $(u^*, u) = 1$ and $xU = u2(u^*, x)$ defines a minimal involution $U \varepsilon \mathfrak{I}$ such that $UP = PU = U$. Since Z commutes with U, we have by Lemma 2.1 that $uZ = u\zeta(u)$, where $\zeta(u) \varepsilon \mathfrak{D}$. Now let v be any other element of $\mathfrak{X}P$. Then, since Z is additive, $(u+v)\zeta(u+v) = u\zeta(u) + v\zeta(v)$ and hence $u(\zeta(u+v) - \zeta(u)) = v(\zeta(v) - \zeta(u+v))$. If u, v are linearly independent, then $\zeta(u+v) = \zeta(u)$ and $\zeta(u+v) = \zeta(v)$ so that $\zeta(u) = \zeta(v)$. If u, v are linearly dependent, choose v' linearly independent of u, v. Then $\zeta(v') = \zeta(u)$ and $\zeta(v') = \zeta(v)$. Therefore $\zeta(u) = \zeta(v)$ in this case as well. In other words $\zeta(u)$ is equal to a constant ζ. That ζ is in the center of $\mathfrak{D}$ follows from the linearity of Z. Now, if Z is an involution, then $Z^2 = Z2$, which implies $\zeta(u) = 2$ or 0, for every $u \varepsilon \mathfrak{X}P$. Therefore, if $\zeta(u) \neq \zeta(v)$, then either $\zeta(u+v) = \zeta(u)$ or $\zeta(u+v) = \zeta(v)$. Since $u(\zeta(u+v) - \zeta(u)) = v(\zeta(v) - \zeta(u+v))$, the first implies $v = 0$ and the second implies $u = 0$. Hence, if u and v are not zero, then $\zeta(u) = \zeta(v)$. This completes the proof.

3.2 Lemma. *If* $\dim(\mathfrak{X}) > 3$, *then ν is equal to either* 16 *or* 8 *according as $I2 \varepsilon \mathfrak{G}$ or $I2 \notin \mathfrak{G}$.*

Let T_1, T_2 be any two commutative involutions in $\mathfrak{I}$ and set $E_i = T_i\frac{1}{2}$ $(i = 1, 2)$. Then E_i, $I - E_i$ are continuous projections on $\mathfrak{X}$. Since E_1, E_2 commute, the following are also continuous projections:

$$P_1 = E_1E_2,\ P_2 = E_1(I - E_2),\ P_3 = (I - E_1)E_2,\ P_4 = (I - E_1)(I - E_2).$$

Also $P_iP_j = 0$ for $i \neq j$. Observe that $c(T_1, T_2)$ consists of all $T \varepsilon \mathfrak{I}$ such that T commutes with each P_i.

Now let U_i denote any minimal involution in $\mathfrak{I}$ such that $P_iU_i = U_iP_i = U_i$. Clearly, for $j \neq i$, $U_iP_j = P_jU_i = 0$ so that $U_i \varepsilon c(T_1, T_2)$. By

Lemma 3.1, if $T \varepsilon c(c(T_1, T_2))$, then $P_iT = P_i\delta_i$, where $\delta_i = 2$ or 0 $(i = 1, 2, 3, 4)$. Since $I = \Sigma P_i$, we have $T = \Sigma P_i\delta_i$. Conversely, any T of this form is an involution and, provided it belongs to $\mathfrak{J}$, is in $c(c(T_1, T_2))$. It follows that $\nu \leq 16$, if $I2 \varepsilon \mathfrak{G}$, and a simple check shows that $\nu \leq 8$, if $I2 \notin \mathfrak{G}$. It remains to prove equality.

Choose linearly independent elements $u_i \varepsilon \mathfrak{X}$ and $u^*_i \varepsilon \mathfrak{X}^*$ such that $(u^*_i, u_j) = \delta_{ij}$ $(i, j = 1, 2, 3)$. Define U_i by $xU_i = u_i2(u^*_i, x)$; then $U_i \varepsilon \mathfrak{J}$ $(i = 1, 2, 3)$. Also define $T_1 = U_1 \circ U_2$ and $T_2 = U_1 \circ U_3$. Then T_1, T_2 commute and the associated projections P_i are given as follows: $P_1 = U_1\frac{1}{2}$, $P_2 = U_2\frac{1}{2}$, $P_3 = U_3\frac{1}{2}$, $P_4 = I - (P_1 + P_2 + P_3)$. Note that $P_1 + P_2 + P_3 = (U_1 \circ U_2 \circ U_3)\frac{1}{2}$. Since $\dim(\mathfrak{X}) > 3$, none of the projections P_i is zero. It follows now that $\nu(T_1, T_2) = 16$, if $I2 \varepsilon \mathfrak{G}$, and $\nu(T_1, T_2) = 8$, if $I2 \notin \mathfrak{G}$.

3.3 THEOREM. *If* $\dim(\mathfrak{X}) \leq 3$, *then every involution on* $\mathfrak{X}$ *is extremal and* $\nu_T = \nu$ *for all* T. *If* $\dim(\mathfrak{X}) > 3$, *then a necessary and sufficient condition for* $T \varepsilon \mathfrak{J}$ *to be extremal is that* $\nu_T = \frac{1}{2}\nu$.

The first statement is obvious. To prove the second, consider an arbitrary $T_1 \varepsilon \mathfrak{J}$ with subspaces $\mathfrak{M}$, $\mathfrak{N}$. Let T_2 be any element of $\mathfrak{J}$ which commutes with T_1 and define P_i (i = 1, 2, 3, 4) as in the proof of Lemma 3.2. If T_1 is minimal, then either $P_1 = 0$ or $P_2 = 0$ and, if T_1 is maximal, then $I2 \varepsilon \mathfrak{G}$ and either $P_3 = 0$ or $P_4 = 0$. In either case, it follows that $\nu(T_1, T_2) \leq \frac{1}{2}\nu$ and hence that $\nu_{T_1} \leq \frac{1}{2}\nu$. Now choose $u_1 \varepsilon \mathfrak{M}$, $u_2 \varepsilon \mathfrak{N}$ and $v^*_i \varepsilon \mathfrak{X}^*$ such that $(v^*_i, u_j) = \delta_{ij}$ $(i, j = 1, 2)$. Define $u^*_1 = v^*_1E^*_1$ and $u^*_2 = v^*_2(I - E_1)^*$, where $E_1 = T_1\frac{1}{2}$. Then $(u^*_i, u_j) = \delta_{ij}$ $(i, j = 1, 2)$. If U_i is defined by $xU_i = u_i2(u^*_i, x)$ $(i = 1, 2)$ and $T_2 = U_1 \circ U_2$, then $T_2 \varepsilon \mathfrak{J}$ and T_1, T_2 commute. With T_2 chosen in this way, since $\dim(\mathfrak{X}) > 3$, at most one of the associated projections P_i can be zero and this happens if, and only if, T_1 is extremal. It follows that $\nu(T_1, T_2) = \nu$ if T_1 is not extremal, and $\nu(T_1, T_2) = \frac{1}{2}\nu$ if T_1 is extremal. This completes the proof.

4. Isomorphic groups of transformations. In this section we consider two systems, $\{\mathfrak{X}, \mathfrak{D}, \mathfrak{X}^* : \mathfrak{G}\}$ and $\{\mathfrak{Y}, \mathcal{E}, \mathfrak{Y}^* : \mathfrak{H}\}$, of the type discussed above, with $\mathfrak{J}$ and $\mathcal{J}$ as the sets of involutions in $\mathfrak{G}$ and $\mathfrak{H}$ respectively. We also assume throughout that the groups $\mathfrak{G}$ and $\mathfrak{H}$ are either isomorphic or anti-isomorphic. The isomorphism or anti-isomorphism of $\mathfrak{G}$ onto $\mathfrak{H}$ will be denoted by $\mathfrak{g} : G \rightarrow \mathfrak{g}(G)$. It is evident that $\mathfrak{g}$ sets up a one-to-one correspondence between the sets $\mathfrak{J}$ and $\mathcal{J}$. Also, $I2 \varepsilon \mathfrak{G}$ if, and only if, $I2 \varepsilon \mathfrak{H}$ and furthermore, if $I2 \varepsilon \mathfrak{G}$, then $\mathfrak{g}(I2) = I2$. By Theorem 3.3, an extremal involution must correspond under $\mathfrak{g}$ to an extremal involution. Moreover, if T_i $(i = 1, \cdots, k)$ are arbitrary extremal involutions in $\mathfrak{I}$, then repeated application of Theorem 2.6, plus the observation that two extremal involutions which commute and have a common subspace must have the same subspaces, shows that the T_i have a subspäce in common if, and only if, the $\mathfrak{g}(T_i)$ do.

Let (T_1, T_2) and (T_3, T_4) be two pairs of non-commutative extremal involutions. Then the pairs are said to be *similar* provided each has a subspace in common and the common subspaces are either both one-dimensional or both non-one-dimensional.

4.1 LEMMA. *The pairs* (T_1, T_2) *and* (T_3, T_4) *are similar if, and only if,* $(\mathfrak{g}(T_1), \mathfrak{g}(T_2))$ *and* $(\mathfrak{g}(T_3), \mathfrak{g}(T_4))$ *are similar.*

Let (T_1, T_2), (T_3, T_4) be similar and assume first that the common subspaces are one-dimensional. If these one-dimensional subspaces coincide, then the T_i $(i = 1, 2, 3, 4)$ have a common subspace. Therefore the $\mathfrak{g}(T_i)$ also have a common subspace and this implies the desired result. If the one-dimensional subspaces are different, let one be generated by u_1 and the other by u_2. Then u_1, u_2 are linearly independent and there exists $u^* \varepsilon \mathfrak{X}^*$ such that $(u^*, u_i) = 1$ $(i = 1, 2)$. Define U_i by $xU_i = u_i 2(u^*, x)$; then $U_i \varepsilon \mathfrak{I}$ $(i = 1, 2)$ and the involutions in each of the triples T_1, T_2, U_1 and T_3, T_4, U_2 have a common subspace. Moreover U_1, U_2 have a common subspace but (U_1, U_2) is not similar to either (T_1, T_2) or (T_3, T_4). Now suppose $(\mathfrak{g}(T_1), \mathfrak{g}(T_2))$ and $(\mathfrak{g}(T_3), \mathfrak{g}(T_4))$ not similar. Then one of these pairs, say the first, must be similar to the pair $(\mathfrak{g}(U_1), \mathfrak{g}(U_2))$. Since $\mathfrak{g}(T_1)$, $\mathfrak{g}(T_2)$, $\mathfrak{g}(U_1)$ have a subspace in common, it follows that $\mathfrak{g}(T_1)$, $\mathfrak{g}(T_2)$, $\mathfrak{g}(U_1)$, $\mathfrak{g}(U_2)$ have a common subspace. This implies in turn that T_1, T_2, U_1, U_2 have a common subspace and contradicts the fact that (T_1, T_2) and (U_1, U_2) are not similar. Therefore $(\mathfrak{g}(T_1), \mathfrak{g}(T_2))$, $(\mathfrak{g}(T_3), \mathfrak{g}(T_4))$ are similar. Finally if the common subspaces in (T_1, T_2) and (T_3, T_4) are not one-dimensional, then the above argument applied to (T^*_1, T^*_2) and (T^*_3, T^*_4) leads again to the desired result and completes proof of the necessity. The sufficiency clearly follows by symmetry.

The one-to-one correspondence between the sets of involutions $\mathfrak{I}$ and $\mathfrak{J}$ will now be used to construct a one-to-one correspondence between the one-dimensional subspaces of $\mathfrak{X}$ and those of either $\mathfrak{Y}$ or $\mathfrak{Y}^*$ such that linear dependence is preserved. The one-dimensional subspace of $\mathfrak{X}$ determined by a non-zero element $x \varepsilon \mathfrak{X}$ will be denoted by $x\mathfrak{D}$. The set of all one-dimensional subspaces of $\mathfrak{X}$ will be denoted by $\mathfrak{X}_{\mathfrak{D}}$. Similar notations will be used for the other spaces under consideration. If $x\mathfrak{D} \varepsilon \mathfrak{X}_{\mathfrak{D}}$ and $\mathfrak{D}x^* \varepsilon \mathfrak{X}^*_{\mathfrak{D}}$, then $x\mathfrak{D}$ and $\mathfrak{D}x^*$ are said to be *orthogonal* provided $(x^*, x) = 0$. If T is an extremal involution in $\mathfrak{J}$, then there exist non-orthogonal $u\mathfrak{D} \varepsilon \mathfrak{X}_{\mathfrak{D}}$ and $\mathfrak{D}u^* \varepsilon \mathfrak{X}^*_{\mathfrak{D}}$ such that $u\mathfrak{D}$ and $(\mathfrak{D}u^*)^{\perp}$ are the subspaces of T. Conversely, every non-orthogonal pair $u\mathfrak{D}$, $\mathfrak{D}u^*$ is associated in this way with an extremal involution. If $I2 \not\varepsilon \mathfrak{S}$, then this association is one-to-one and T is minimal. If $I2 \varepsilon \mathfrak{S}$, then both T and $I2 - T$ are associated with $u\mathfrak{D}$ and $\mathfrak{D}u^*$.

4. 2 THEOREM. *There exist two one-to-one mappings F and F^* such that F maps $\mathfrak{X}_{\mathfrak{D}}$ onto either $\mathfrak{Y}_{\mathcal{E}}$ or $\mathfrak{Y}^*_{\mathcal{E}}$ while F^* maps $\mathfrak{X}^*_{\mathfrak{D}}$ onto either $\mathfrak{Y}^*_{\mathcal{E}}$ or $\mathfrak{Y}_{\mathcal{E}}$. Also $(x_3\mathfrak{D})F \subseteq (x_1\mathfrak{D})F + (x_2\mathfrak{D})F$ or $(\mathfrak{D}^*x_3)F^* \subseteq (\mathfrak{D}x^*_1)F^* + (\mathfrak{D}x^*_2)F^*$ if, and only if, $x_3\mathfrak{D} \subseteq x_1\mathfrak{D} + x_2\mathfrak{D}$ or $\mathfrak{D}x^*_3 \subseteq \mathfrak{D}x^*_1 + \mathfrak{D}x^*_2$ respectively.*

Since $\mathfrak{g}(T)$ is extremal if, and only if, T is extremal, it is evident that $\mathfrak{g}$ establishes a one-to-one correspondence $\{\mathfrak{D}x^*, x\mathfrak{D}\} \leftrightarrow \{\mathcal{E}y^*, y\mathcal{E}\}$ between non-orthogonal pairs from $\mathfrak{X}^*_{\mathfrak{D}} \times \mathfrak{X}_{\mathfrak{D}}$ and $\mathfrak{Y}^*_{\mathcal{E}} \times \mathfrak{Y}_{\mathcal{E}}$. Now let $\{\mathfrak{D}x^*, x\mathfrak{D}\}$ be an arbitrary non-orthogonal pair from $\mathfrak{X}^*_{\mathfrak{D}} \times \mathfrak{X}_{\mathfrak{D}}$ and let $\mathfrak{D}x^*_1$ be any element of $\mathfrak{X}^*_{\mathfrak{D}}$ distinct from $\mathfrak{D}x^*$ and also non-orthogonal to $x\mathfrak{D}$. Then $\{\mathfrak{D}x^*, x\mathfrak{D}\} \leftrightarrow \{\mathcal{E}y^*, y\mathcal{E}\}$ and $\{\mathfrak{D}x^*_1, x\mathfrak{D}\} \leftrightarrow \{\mathcal{E}y^*_1, y_1\mathcal{E}\}$. By Theorem 2. 6, either $y_1\mathcal{E} = y\mathcal{E}$ or $\mathcal{E}y^*_1 = \mathcal{E}y^*$ and, by Lemma 4. 1, whichever case holds is independent of the choice of $x\mathfrak{D}$, $\mathfrak{D}x^*$ and $\mathfrak{D}x^*_1$. Therefore we can define, without ambiguity, $(x\mathfrak{D})F = y\mathcal{E}$ if $y_1\mathcal{E} = y\mathcal{E}$, and $(x\mathfrak{D})F = \mathcal{E}y^*$ if $\mathcal{E}y^*_1 = \mathcal{E}y^*$. In the first case F maps $\mathfrak{X}_{\mathfrak{D}}$ onto $\mathfrak{Y}_{\mathcal{E}}$, and in the second it maps $\mathfrak{X}_{\mathfrak{D}}$ onto $\mathfrak{Y}^*_{\mathcal{E}}$. The mapping F^* is defined in an exactly similar way. It is not difficult to verify that F and F^* are one-to-one. Observe that $\{\mathfrak{D}x^*, x\mathfrak{D}\} \leftrightarrow \{(\mathfrak{D}x^*)F^*, (x\mathfrak{D})F\}$ or $\{\mathfrak{D}x^*, x\mathfrak{D}\} \leftrightarrow \{(x\mathfrak{D})F, (\mathfrak{J}x^*)F^*\}$ according as F maps $\mathfrak{X}_{\mathfrak{D}}$ onto $\mathfrak{Y}_{\mathcal{E}}$ or $\mathfrak{Y}^*_{\mathcal{E}}$. It remains to prove the last statement of the theorem. It will be sufficient to consider only F and prove that $x_3\mathfrak{D} \subseteq x_1\mathfrak{D} + x_2\mathfrak{D}$ implies $(x_3\mathfrak{D})F \subseteq (x_1\mathfrak{D})F + (x_2\mathfrak{D})F$. If $x_3\mathfrak{D}$ is equal to either $x_1\mathfrak{D}$ or $x_2\mathfrak{D}$, the desired result is immediate; therefore assume $x_3\mathfrak{D} \neq x_1\mathfrak{D}$ and $x_3\mathfrak{D} \neq x_2\mathfrak{D}$. Then also $x_1\mathfrak{D} \neq x_2\mathfrak{D}$. Choose $x_i \varepsilon x_i\mathfrak{D}$ so that $x_1 + x_2 \varepsilon x_3\mathfrak{D}$ and $x_3 = (x_1 + x_2)\frac{1}{2}$. Choose $x^* \varepsilon \mathfrak{X}^*$ such that $(x^*, x_1) = (x^*, x_2) = 1$; then also $(x^*, x_3) = 1$. Define involutions $X_i \varepsilon \mathfrak{J}$ as follows: $xX_i = x_i 2(x^*, x)$ $(i = 1, 2, 3)$. Observe that the subspaces of X_i are $x_i\mathfrak{D}$ and $(\mathfrak{D}x^*)^{\perp}$. Since $x_3\mathfrak{D} \subseteq x_1\mathfrak{D} + x_2\mathfrak{D}$, it follows, by Lemma 2, 5, that $X_3 \varepsilon c(c(X_1, X_2))$. Hence also $\mathfrak{g}(X_3) \varepsilon c(c(\mathfrak{g}(X_1), \mathfrak{g}(X_2)))$ and $\mathfrak{g}(X_3)^* \varepsilon c(c(\mathfrak{g}(X_1)^*, \mathfrak{g}(X_2)^*))$. Let $y_i\mathcal{E}$ and $(\mathcal{E}y^*_i)^{\perp}$ denote the subspaces of $\mathfrak{g}(X_i)$ $(i = 1, 2, 3)$. It follows from Lemma 2. 5 that $y_3\mathcal{E} \subseteq y_1\mathcal{E} + y_2\mathcal{E}$. Similarly, by replacing $\mathfrak{g}(X_i)$ by $g(X_i)^*$, we obtain $\mathcal{E}y^*_3 \subseteq \mathcal{E}y^*_1$

$+ \mathcal{E}y^*_2$. Since $(x_i\mathcal{D})F = y_i\mathcal{E}$ or $(x_i\mathcal{D})F = \mathcal{E}y^*_i$, according as F maps $\mathfrak{X}_{\mathcal{D}}$ onto $\mathfrak{Y}_{\mathcal{E}}$ or $\mathfrak{Y}^*_{\mathcal{E}}$, the proof is complete.

4. 3 LEMMA. *Let $\mathfrak{M}$, $\mathfrak{N}$ be the subspaces of an involution $T \varepsilon \mathfrak{J}$. Then the linear subspaces determined by $(\mathfrak{M}_{\mathcal{D}})F$ and $(\mathfrak{N}_{\mathcal{D}})F$ in either $\mathfrak{Y}$ or $\mathfrak{Y}^*$ are the subspaces of the involution $\mathfrak{g}(T)$ or $\mathfrak{g}(T)^*$ according as F maps $\mathfrak{X}_{\mathcal{D}}$ onto $\mathfrak{Y}_{\mathcal{E}}$ or $\mathfrak{Y}^*_{\mathcal{E}}$.*

It will be enough to consider the case in which F maps $\mathfrak{X}_{\mathcal{D}}$ onto $\mathfrak{Y}_{\mathcal{E}}$. Let $\mathfrak{M}'$, $\mathfrak{N}'$, be the subspaces of $\mathfrak{g}(T)$ and consider the subspace $\mathfrak{M}$ of T. If $\mathfrak{M}$ is one-dimensional, then from the definition of F, we already know that $(\mathfrak{M}_{\mathcal{D}})F$ is equal to either $\mathfrak{M}'_{\mathcal{E}}$ or $\mathfrak{N}'_{\mathcal{E}}$. If $\mathfrak{M}$ is not one-dimensional, let x_1, x_2 be any two linearly independent elements of $\mathfrak{M}$ and choose $v^* \varepsilon \mathfrak{X}^*$ such that $(v^*, x_i) = 1$ $(i = 1, 2)$. Let $u^* = \frac{1}{2}v^*T^*$, then still $(u^*, x_i) = 1$ and also $\mathfrak{N} \subseteq (\mathcal{D}u^*)^{\perp}$. Define the involutions X_i by $xX_i = x_i2(u^*, x)$ $(i = 1, 2)$. Then, by Lemma 2. 2, $T \varepsilon c(X_1, X_2)$. Hence $\mathfrak{g}(T) \varepsilon c(\mathfrak{g}(X_1), \mathfrak{g}(X_2))$ and, by Lemma 2. 4, either $\mathfrak{M}'$ or $\mathfrak{N}'$ must contain $(x_1\mathcal{D})F + (x_2\mathcal{D})F$. By holding $x_1 \varepsilon \mathfrak{M}$ fixed and varying x_2 over $\mathfrak{M}$, we conclude that either $\mathfrak{M}'_{\mathcal{E}}$ or $\mathfrak{N}'_{\mathcal{E}}$, say the first, must contain $(\mathfrak{M}_{\mathcal{D}})F$. It follows easily by symmetry that $\mathfrak{M}'_{\mathcal{E}} = (\mathfrak{M}_{\mathcal{D}})F$. A similar argument gives $\mathfrak{N}'_{\mathcal{E}} = (\mathfrak{N}_{\mathcal{D}})F$ and completes the proof.

The next lemma appears to be well-known in modern projective geometry. However, since we do not have a reference for it, we will include a brief outline of a proof.

4. 4 LEMMA. *Let $\mathfrak{X}$ be any right $\mathcal{D}$-space with* $\dim(\mathfrak{X}) \geqq 3$ *and let $\mathfrak{Y}$ be a right or left $\mathcal{E}$-space. Also let F be any one-to-one mapping of $\mathfrak{X}_{\mathcal{D}}$ onto $\mathfrak{Y}_{\mathcal{E}}$ such that $(x_3\mathcal{D})F \subseteq (x_1\mathcal{D})F + (x_2\mathcal{D})F$ if, and only if, $x_3\mathcal{D} \subseteq x_1\mathcal{D} \dotplus x_2\mathcal{D}$. Then there exists an isomorphic mapping Φ of $\mathfrak{X}$ onto $\mathfrak{Y}$ such that $(x\mathcal{D})\Phi = (x\mathcal{D})F$.*

It will be sufficient to take $\mathfrak{Y}$ as a right $\mathcal{E}$-space. Let x_0 be a fixed non-zero element of $\mathfrak{X}$ and denote by y_0 an arbitrary but fixed non-zero element of $(x_0\mathcal{D})F$. We define $0\Phi_0 = 0$ and $x_0\Phi_0 = y_0$. If x is linearly independent of x_0, then x, $x_0 - x$ are linearly independent so that $(x\mathcal{D})F \neq ((x_0 - x)\mathcal{D})F$. Moreover, since $x_0\mathcal{D} \subseteq x\mathcal{D} + (x_0 - x)\mathcal{D}$, we have $(x_0\mathcal{D})F \subseteq (x\mathcal{D})F + ((x_0 - x)\mathcal{D})F$. Therefore there exist uniquely determined elements $y \varepsilon (x\mathcal{D})F$ and $y' \varepsilon ((x_0 - x)\mathcal{D})F$ such that $y_0 = y + y'$. Set $x\Phi_0 = y$. This defines Φ_0 for 0, x_0 and all x linearly independent of x_0. Now consider a second fixed element x_1 linearly independent of x_0 and define Φ_1 exactly as Φ_0 was defined except with x_0 replaced by x_1 and $x_1\Phi_1 = x_1\Phi_0$.

It turns out that $x\Phi_1 = x\Phi_0$ whenever both are defined. If $x\Phi = x\Phi_0$ or $x\Phi_1$, whichever is defined, then Φ is single-valued and can be shown to be additive. Now, for $\lambda \varepsilon \mathfrak{D}$, $(x_0\lambda)\Phi \varepsilon (x_0\mathfrak{D})F$. Therefore, there exists $\lambda^\phi \varepsilon \mathcal{E}$ such that $(x_0\lambda)\Phi = x_0\Phi\lambda^\phi$. Moreover it can be proved that $(x\lambda)\Phi = x\Phi\lambda^\phi$, for every x, and that $\lambda \rightarrow \lambda^\phi$ is an isomorphism of $\mathfrak{D}$ onto $\mathcal{E}$. Therefore Φ maps $\mathfrak{X}$ isomorphically onto $\mathfrak{Y}$ and it is obvious that $(x\mathfrak{D})\Phi = (x\mathfrak{D})F$.

4.5 THEOREM. *There exist one-to-one mappings* Φ, Ψ *such that one of the following two cases is true*:

(i) Φ *and* Ψ *are continuous isomorphisms of* $\mathfrak{X}$ *onto* $\mathfrak{Y}$ *and* $\mathfrak{X}^*$ *onto* $\mathfrak{Y}^*$ *respectively, involving the same isomorphism* $\lambda \rightarrow \lambda^\phi$ *of* $\mathfrak{D}$ *onto* $\mathcal{E}$ *and such that* $(x^*, y\Phi^{-1})^\phi \equiv (x^*\Psi, y)$.

(ii) Φ *and* Ψ *are continuous isomorphisms of* $\mathfrak{X}$ *onto* $\mathfrak{Y}^*$ *and* $\mathfrak{X}^*$ *onto* $\mathfrak{Y}$ *respectively, involving the same anti-isomorphism* $\lambda \rightarrow \lambda^\phi$ *of* $\mathfrak{D}$ *onto* $\mathcal{E}$ *and such that* $(y\Psi^{-1}, x)^\phi \equiv (x\Phi, y)$.

In both cases $(x\mathfrak{D})\Phi = (x\mathfrak{D})F$ *and* $(\mathfrak{D}x^*)\Psi = (\mathfrak{D}x^*)F^*$, *where* F, F^* *are the mappings given in Theorem* 4.2.

We will discuss only case (i) which corresponds to the case in which the mapping F of Theorem 4.2 takes $\mathfrak{X}_{\mathfrak{D}}$ onto $\mathfrak{Y}_{\mathcal{E}}$. Denote by Φ an isomorphism of $\mathfrak{X}$ onto $\mathfrak{Y}$ given by Lemma 4.4. By Lemma 4.3, $(\mathfrak{D}x^*)F^* = \mathcal{E}y^*$ if, and only if, $(\mathfrak{D}x^*)\Phi = \mathcal{E}y^*$. It follows from this that, for fixed $x^* \varepsilon \mathfrak{X}^*$, $(x^*, y\Phi^{-1})^\phi$ defines an element of $(\mathfrak{D}x^*)F^*$ which we denote by $x^*\Psi$. It is not difficult to show that Ψ has the desired properties and also that both Φ, Ψ are continuous.

5. Representation of the group isomorphism.

We turn in this section to the problem of representing $G \rightarrow \mathfrak{g}(G)$ in terms of the isomorphism of the underlying vector spaces given by Theorem 4.5. It will be assumed throughout that $\mathfrak{g}$ is an isomorphism.

Consider the set of elements of the center of the division ring $\mathcal{E}$ minus the identity element 1. This set is a group under the circle operation and, as such, will be denoted by $\mathcal{E}_0$.

The following theorem provides the desired representation of the group isomorphism $\mathfrak{g}$. For the finite dimensional case, it reduces essentially to a theorem of Dieudonné [1, Théorème 1].

5.1 THEOREM. *There exists a homomorphism* $G \rightarrow \gamma(G)$ *of* $\mathfrak{G}$ *into* $\mathcal{E}_0$ *such that one of the following is true for all* $G \varepsilon \mathfrak{G}$:

(i) $\mathfrak{g}(G) = (\Phi^{-1}G\Phi) \circ I_{\gamma}(G)$; (ii) $\mathfrak{g}(G) = (\Phi^{-1}G^{\circ}\Phi)^{*} \circ I_{\gamma}(G)$,

where Φ *is the isomorphism given in the corresponding case of Theorem* 4. 5.

Consider first case (i) in which Φ is an isomorphism of $\mathfrak{X}$ onto $\mathfrak{Y}$. Let T be a minimal involution in $\mathfrak{G}$ with subspaces $\mathfrak{M}$ and $\mathfrak{N}$. Then, by Lemma 4. 3, the involution $\mathfrak{g}(T)$ has $(\mathfrak{M})\Phi$ and $(\mathfrak{N})\Phi$ as its subspaces. On the other hand, $\Phi^{-1}T\Phi$ is also an involution with subspaces $(\mathfrak{M})\Phi$ and $(\mathfrak{N})\Phi$. It follows that $\mathfrak{g}(T) = (\Phi^{-1}T\mathfrak{N}) \circ I_{\rho}$, where $\rho = 0$ or 2. Therefore $\mathfrak{g}(T)$ has the form (i) for minimal involutions.

We now define the mapping $G \to G^{\sigma} = \Phi\mathfrak{g}(G)\Phi^{-1}$, which is an isomorphism of the group $\mathfrak{G}$ onto another group of linear transformations in $\mathfrak{X}$. Observe that, if T is a minimal involution, then $T^{\sigma} = T \circ I_{\rho}$, where $\rho = 0$ or 2.

Consider an arbitrary $G \varepsilon \mathfrak{G}$ and an arbitrary minimal involution $T \varepsilon \mathfrak{G}$. Then $G \circ T \circ G^{\circ}$ is also a minimal involution. In fact, $G \circ T \circ G^{\circ} = (I - G)T(I - G)^{-1}$, where $(I - G)^{-1}$ is the ordinary multiplicative inverse of $I - G$. Therefore it follows that $(G \circ T \circ G^{\circ})^{\sigma} = (G \circ T \circ G^{\circ}) \circ I_{\rho_1}$, where $\rho_1 = 0$ or 2. But $(G \circ T \circ G^{\circ})^{\sigma} = G^{\sigma} \circ T^{\sigma} \circ (G^{\sigma})^{\circ}$ and $T^{\sigma} = T \circ I_{\rho_2}$, where $\rho_2 = 0$ or 2. Hence $G^{\sigma} \circ T \circ (G^{\sigma})^{\circ} = (G \circ T \circ G^{\circ}) \circ I_{\rho}$, where $\rho = \rho_1 \circ \rho_2 = 0$ or 2. This last equation can be written in the form $I\rho = (I - G^{\sigma})T(I - G^{\sigma})^{-1} + (I - G)T(I - G)^{-1}(\rho - 1)$. Each term on the right is one-dimensional and consequently the right hand side is at most two-dimensional. Since $\mathfrak{X}$ is at least three-dimensional, we conclude that $\rho = 0$. In other words,

$$(1) \qquad (I - G^{\sigma})T(I - G^{\sigma})^{-1} = (I - G)T(I - G)^{-1}.$$

Suppose now that there exists an $x \varepsilon \mathfrak{X}$ such that the vectors $u = x(I - G)$ and $v = x(I - G^{\sigma})$ are linearly independent. We can then choose T such that $uT = u2$ and $vT = 0$. An application of equation (1) to the vector x then yields $u2(I - G)^{-1} = 0$. This obviously implies $u = 0$, a contradiction. It follows that, for every $x \varepsilon \mathfrak{X}$, the vectors $x(I - G^{\circ})$ and $x(I - G)$ are linearly dependent. An argument similar to the above also shows that these vectors vanish simultaneously. Therefore, for every $x \varepsilon \mathfrak{X}$, there exists $\lambda_x \varepsilon \mathfrak{D}$ such that $x(I - G^{\sigma}) = x(I - G)\lambda_x$. Hence $x(I - G^{\sigma})(I - G)^{-1} = x\lambda_x$. By additivity, we obtain $x(\lambda_x - \lambda_{x+y}) + y(\lambda_y - \lambda_{x+y}) = 0$, for all $x, y \varepsilon \mathfrak{X}$. If x, y are linearly independent, this implies $\lambda_x = \lambda_{x+y} = \lambda_y$. If x, y are linearly dependent, choose $z \varepsilon \mathfrak{X}$ linearly independent of x, y. Then $\lambda_x = \lambda_z = \lambda_y$ so that $\lambda_x = \lambda_y$ in this case as well. Therefore λ_x is equal to a constant λ independent of x. In other words, $I - G^{\sigma} = (I - G)\lambda$, and this can be written in the form $G^\sigma = G \circ I(1-\lambda)$. It is immediate from linearity that λ is in the center of $\mathfrak{D}$. Also, since $I(1-\lambda) = G^\circ \circ G^\sigma$ and $(G^\circ \circ G^\sigma)^\circ$ exists, it follows that $1-\lambda$ is in $\mathfrak{D}_0$. Returning to $\mathfrak{g}(G)$, we have $\mathfrak{g}(G) = (\Phi^{-1}G\Phi) \circ I\gamma(G)$, where $\gamma(G) = (1-\lambda)^\Phi$ and $\lambda \rightarrow \lambda^\Phi$ is the isomorphism of $\mathfrak{D}$ onto $\mathcal{E}$ associated wth Φ. It is obvious that $G \rightarrow \gamma(G)$ must be a homomorphism of $\mathfrak{G}$ into $\mathcal{E}_0$. This completes the proof for case (i).

In case (ii) we consider $G \rightarrow G^\delta = \Phi\mathfrak{g}(G)^*\Phi^{-1}$ instead of $G \rightarrow G^\sigma$. Observe that $G \rightarrow G^\delta$ is an anti-isomorphism of $\mathfrak{G}$. An argument similar to that given in case (i) yields, in place of equation (1),

$$(1)^* \qquad (I-G^\delta)^{-1}T(I-G^\delta) = (I-G)T(I-G)^{-1}.$$

Again as in case. (i), we obtain that $I-G = (I-G^\delta)^{-1}\lambda$, where $\lambda \varepsilon \mathfrak{D}_0$. This can be written in the form $G^\delta = G^\circ \circ I(1-\lambda)$. Therefore

$$\mathfrak{g}(G) = (\Phi^{-1}G^\circ\Phi)^* \circ I\gamma(G),$$

where $\gamma(G) = (1-\lambda)^\Phi$. This completes the proof.

YALE UNIVERSITY.

REFERENCES.

1. J. Dieudonné, "Sur les automorphismes des groupes classiques," *Comptes Rendus de l'Académie des Sciences de Paris*, vol. 225 (1947), pp. 914-915.
2. E. Hille, *Functional analysis and semigroups.* American Mathematical Society Colloquium Publications, vol. 31, New York (1948).
3. N. Jacobson, "On the theory of primitive rings," *Annals of Mathematics*, vol. 48 (1947), pp. 8-21.
4. I. Kaplansky, "Topological rings," *American Journal of Mathematics*, vol. 59 (1947), pp. 153-183.
5. G. Mackey, "Isomorphisms of normed linear spaces," *Annals of Mathematics*, vol. 43 (1942), pp. 244-260.
6. O. Schreier and B. L. van der Waerden, "Die Automorphismen der projektiven Gruppen," *Abhandlungen des mathematisches Seminars der Hamburgischen Universität*, vol. 6 (1928), pp. 303-322.

STEADY, ROTATIONAL, PLANE FLOW OF A GAS.*[1]

By Monroe H. Martin.

---

1. Introduction. If one excludes flows for which the stream lines are isobars, the network formed by the stream lines and isobars in the region of the (x, y)-plane covered by the flow can be used as a system of curvilinear coordinates. This amounts to taking the pressure p and stream function ψ for independent variables and to regard the other flow variables, i. e., the velocity components u, v, the density ρ, and the coordinates x, y, as unknown functions of them.

This choice of independent variables has the following advantage. Taking the equation of state of the gas to be $\rho = f(p, S)$, where S denotes the entropy, it is well known that S is constant along a stream line. Consequently, if we specify the variation of entropy from stream line to stream line by setting $S = S(\psi)$, the density ρ becomes a known function $\rho = \rho(p, \psi)$ of the independent variables p, ψ.

The problem of determining the remaining unknown functions x, y, u, v of p, ψ is reduced in Section 4 to the integration of a quasi-linear partial differential equation for a single unknown function $\theta = \arctan v/u = \theta(p, \psi)$, provided $\theta_\psi \not\equiv 0$, that is, provided the flow does not cross each isobar with a constant direction. These exceptional flows are studied in Section 5.

The quasi-linear partial differential equation for θ involves the speed $q = q(p, \psi)$ of the flow, which becomes a known function of p, ψ, once $\rho(p, \psi)$ is given and two arbitrary functions $p_0 = p_0(\psi)$, $q_0 = q_0(\psi)$ are specified. It is, however possible to adopt an inverse approach, and assign θ to be a definite function of p, ψ. This leads to a *linear partial differential equation of the third order* for q. As an example of the inverse method, in Section 6 we study the simplest case, $\theta = \theta(\psi)$, in which the stream lines lie on straight lines.

2. The equations of motion. The dynamical equations and the equations of continuity in steady, plane flow of a fluid subject to no external forces constitute an underdetermined system

---

* Received September 21, 1949.

[1] Sponsored by the Office of Naval Research. Presented April 2, 1949 at the Durham, North Carolina meeting of the American Mathematical Society. This investigation was carried out under project NOL 139 of the Naval Ordnance Laboratory, White Oak, Maryland.

$$\rho(u_x u + u_y v) + p_x = 0, \qquad \rho(v_x u + v_y v) + p_y = 0, \tag{1}$$
$$(\rho u)_x + (\rho v)_y = 0,$$

of three partial differential equations for four unknown functions u, v, ρ, p of x, y. Following the usual notation, u, v designate the *components of the velocity,* ρ the *density,* and p the *pressure* of the fluid at the point (x, y).

Writing this system in the alternative form

$$(p + \rho u^2)_x + (\rho uv)_y = 0, \qquad (\rho uv)_x + (p + \rho v^2)_y = 0,$$
$$(\rho u)_x + (\rho v)_y = 0,$$

it is easy to see that, given any solution

$$u = u(x, y), \qquad v = v(x, y), \qquad \rho = \rho(x, y) \neq 0, \qquad p = p(x, y), \tag{2}$$

of (1), there exist three functions[2] $\bar{\xi}, \bar{\eta}, \psi$ of x, y defined by

$$d\bar{\xi} = -\rho uv dx + (p + \rho u^2) dy, \qquad d\bar{\eta} = -(p + \rho v^2) dx + \rho uv dy, \tag{3}$$
$$d\psi = -\rho v dx + \rho u dy.$$

This differential system is replaced by the system

$$d\xi = -y dp + u d\psi, \qquad d\eta = x dp + v d\psi, \qquad d\psi = -\rho v dx + \rho u dy, \tag{4}$$

when we set $\xi = \bar{\xi} - py$, $\eta = \bar{\eta} + px$.

The form of the first two equations in (4) suggests that p, ψ be employed as independent variables, that is, that the stream lines and isobars be taken as a system of curvilinear coordinates in the region of the (x, y)-plane covered by the flow. We accordingly assume that

$$0 < |J^{-1}| < +\infty, \qquad J^{-1} = \partial(p, \psi)/\partial(x, y), \tag{5}$$

holds throughout the above region. This automatically rules out flows for which the pressure is constant everywhere or for which $p = p(\psi)$, i. e., flows for which the stream lines are isobars. Such flows form a very restricted class[3] and will be excluded from now on. When p, ψ are taken for independent variables, (4) leads to an underdetermined system of four partial differential equations

$$u_p = -y_\psi, \qquad v_p = x_\psi, \qquad uy_p - vx_p = 0; \qquad uy_\psi - vx_\psi = \rho^{-1}, \tag{6}$$

for five unknown functions x, y, u, v, ρ of p, ψ. The first two equations represent the integrability conditions for the first two equations in (4). The last two equations arise from the third equation of (4) when one sets $dx = x_p dp + x_\psi d\psi$, $dy = y_p dp + y_\psi d\psi$, and recalls that p, ψ are independent variables.

---

[2] ψ is, of course, the stream function. For the physical significance of $\bar{\xi}, \bar{\eta}$ see a Technical Memorandum, *A new approach to problems in two-dimensional flow.* NOLM 9869, Naval Ordnance Laboratory, by the author. To appear in Quarterly of Applied Mathematics.

[3] For a treatment of flows for which $p = p(\psi)$ see the memorandum cited in 2.

Starting with a solution (2) of the underdetermined system (1), one uses the last equation in (3) to calculate $\psi = \psi(x, y)$. The pair of equations $p = p(x, y)$, $\psi = \psi(x, y)$, determine x, y as functions of p, ψ, in view of (5). It follows from (2) that u, v, ρ are known functions of p, ψ. Thus, given any solution (2) of the underdetermined system (1), we are led to a solution

$$(7) \quad x = x(p,\psi),\ y = y(p,\psi),\ \ u = u(p,\psi),\ \ v = v(p,\psi),\ \ \rho = \rho(p,\psi) \neq 0,$$

of the underdetermined system (6).

Conversely, given a solution (7) of (6), a solution (2) of (1) can be constructed, provided

$$(8) \qquad 0 < |J| < +\infty, \qquad J = \partial(x,y)/\partial(p,\psi).$$

In view of (8), the first two equations in (7) can be solved for p, ψ. This yields p, ψ and consequently u, v, ρ as functions of x, y. One has, of course, $J\psi_x = -y_p$, $J\psi_y = x_p$, $Jp_x = y_\psi$, $Jp_y = -x_\psi$, and therefore

$$(9) \quad \begin{aligned} Ju_x &= -u_\psi y_p + u_p y_\psi, & Jv_x &= -v_\psi y_p + v_p y_\psi, & J\rho_x &= -\rho_\psi y_p + \rho_p y_\psi, \\ Ju_y &= u_\psi x_p - u_p x_\psi, & Jv_y &= v_\psi x_p - v_p x_\psi, & J\rho_y &= \rho_\psi x_p - \rho_p x_\psi. \end{aligned}$$

When the partial derivatives of u, v, ρ, p with respect to x, y obtained from these equations are substituted in the system (1), it will be found that the system is satisfied by virtue of (6) and the further relation $\rho_p = \rho^2(u_\psi y_p - u_p y_\psi + v_p x_\psi - v_\psi x_p)$. This relation is needed to verify that the continuity equation is satisfied and is obtained from the second column of (6). The first equation is differentiated with respect to ψ, the second with respect to p and the results are subtracted.

To examine the significance of (8), we find from (6) that

$$(10) \qquad uu_p + vv_p = -\rho^{-1},$$

and that $u_p x_p + v_p y_p = -J$, $vx_p - uy_p = 0$. Solving these equations simultaneously for x_p, y_p, and using (10), one finds

$$(11) \qquad x_p = \rho J u, \qquad y_p = \rho J v.$$

It follows that

$$(11') \qquad dp/ds = (\rho J q)^{-1}, \text{ if } q = (u^2 + v^2)^{\frac{1}{2}} > 0,$$

the arc length s being measured along a stream line from a fixed point so that it increases in the direction of flow on the stream line. Thus, provided we assume that the density ρ and speed q remain finite and different from zero, the two conditions $0 < |dp/ds| < +\infty$, $0 < |J| < +\infty$, are equivalent.

The above results are summed up in the theorem below.

THEOREM 1. *The solution of the underdetermined systems* (1) *and* (6) *are equivalent problems, provided the stream lines are not isobars, the density,* ρ, *the speed* q, *and the pressure gradient* dp/ds *along a stream line remain finite and never vanish.*

The system (6) becomes determinate when the equation of state of the gas and the distribution of entropy S on the stream lines is specified by setting $\rho = f(p, S)$, $S = S(\psi)$.

For a polytropic gas, for example, one places

$$\rho = kp^{1/\gamma}, \qquad k = e^{-(S-S_0)/c_p}, \qquad S = S(\psi), \gamma > 1, \tag{12}$$

the function $S(\psi)$ being taken arbitrarily. After the equation of state of the gas and the distribution of entropy on the stream lines has been fixed, the density ρ becomes a known function

$$\rho = \rho(p, \psi), \tag{13}$$

which we term the *density function.* Once it has been assigned, (6) becomes a determinate system of four partial differential equations for four unknown functions x, y, u, v of p, ψ.

A solution

$$\begin{aligned} x &= x(p, \psi), & u &= u(p, \psi), \\ y &= y(p, \psi), & v &= v(p, \psi), \end{aligned} \tag{14}$$

of a determinate system (6) offers a parametric representation of the stream lines and their hodographs in the first and second columns respectively, the pressure p serving for parameter.

The functions in (14) are not necessarily single valued, as it is possible for a point in the (p, ψ)-plane to correspond to more than one point in the physical plane. Such a situation will always arise when an isobar cuts a fixed stream line in more than one point and occurs even in the simplest examples, e. g., the flow[4] defined by the conformal mapping $z = w^2$ in which the stream lines are confocal parabolas and the isobars are circles concentric at the common focus.

Moreover it is also possible for two different points in the (p, ψ)-plane to map into the same point of the physical plane. This physical impossibility may be avoided by confining the point (p,ψ) to a properly chosen region in the (p,ψ)-plane.

---

[4] As the referee has pointed out, another example of great practical importance occurs in the neighborhood of the stagnation point in steady flow about a symmetrical blunt profile.

3. The Bernoulli function. If (10) be integrated between the limits p_0 and p for a fixed ψ, we obtain Bernoulli's theorem

$$\tfrac{1}{2}q^2 = \tfrac{1}{2}q_0^2 - \int_{p_0}^{p} \rho^{-1}dp, \qquad p_0 = p_0(\psi),\, q_0 = q_0(\psi), \tag{15}$$

where the arbitrary functions $p_0(\psi)$, $q_0(\psi)$ are at our disposal. To understand their significance, let us imagine the region covered by the flow referred to its stream lines and isobars as a system of curvilinear coordinates, as in the figure below.

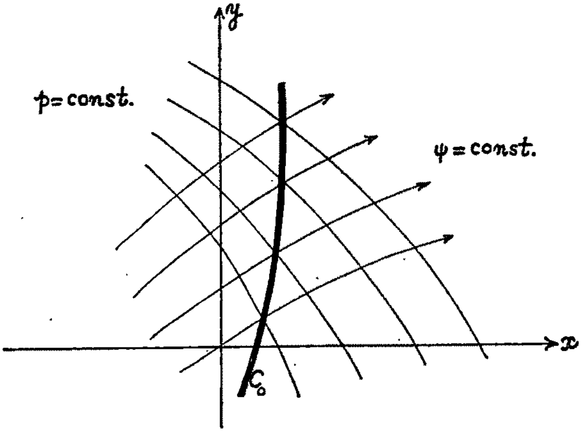


Figure 1.

The choice of $p_0(\psi)$ selects a curve $p = p_0(\psi)$ (the curve C_0) along which the speed of flow is determined by specifying $q_0(\psi)$. In particular, if $p_0 = \text{const.}$, C_0 is an isobar.

For a polytropic gas we find, from (12), that (15) yields

$$q^2 = q_0^2 - 2(p^{1-n} - p_0^{1-n})/\{k(1-n)\}, \qquad 0 < n = 1/\gamma < 1. \tag{16}$$

Taking $p_0 = 0$, this reduces to

$$q^2 = q_0^2 - 2p^{1-n}/\{k(1-n)\}, \tag{17}$$

and C_0 becomes the isobar of zero pressure along which the speed attains its maximum value q_0 (the ultimate speed). If in addition, $q_0 = \text{const.}$, the flow is iso-energetic.

Returning to the general case, we assume that the density function (13) and the arbitrary functions $p_0(\psi)$, $q_0(\psi)$ are specified. From (15) we see

$$q = q(p, \psi) \tag{18}$$

is a known function of p, ψ. This function is termed the *Bernoulli function* of the flow.

Conversely, if the Bernoulli function is given, the density function is uniquely determined by

$$\rho = -(qq_p)^{-1}, \tag{19}$$

provided, of course, that $qq_p < 0$.

An interesting connection between the Bernoulli function q and the vorticity $\omega = u_y - v_x$ follows[5] from (9). In substituting for u_y, v_x and then using (6), (11) to substitute for u_p, v_p, x_p, y_p, one has

$$\omega = \rho q q_\psi = -q_\psi / q_p, \tag{20}$$

the latter equation being a consequence of (19). This formula leads immediately to the following theorem:

THEOREM 2. *Under the hypotheses of Theorem* 1, *the flow will be irrotational if, and only if, the Bernoulli function is independent of* ψ, *that is* $q = q(p)$.

Differentiating (19) with respect to p and introducing the Mach number $M = q/G$, where G denotes the local velocity of sound, one finds that $M^2 - 1 = \rho^2 q^3 q_{pp}$. As a consequence, we have the theorem:

THEOREM 3. *Under the hypotheses of Theorem* 1, *the flow is subsonic, or supersonic, according as the Bernoulli function satisfies the inequality* $q_{pp} < 0$, *or the inequality* $q_{pp} > 0$.

As a corollary we note that if q is a linear function of p, the flow is always sonic.

4. The quasi-linear partial differential equation.

The integration of a determinate system (6) may be reduced to the integration of a quasi-linear partial differential equation.

Let (14) be a solution of (6) and let $q = (u^2 + v^2)^{\frac{1}{2}} = q(p, \psi)$ be the associated Bernoulli function. If we introduce the angle of inclination θ of the stream lines by setting

---

[5] For a different proof see 2.

$$u = q\cos\theta, \qquad v = q\sin\theta, \tag{21}$$

it follows from (11) that $\theta = \arctan(y_p/x_p)$ and, from (11′), that

$$x_p = (dp/ds)^{-1}\cos\theta, \qquad y_p = (dp/ds)^{-1}\sin\theta. \tag{22}$$

Moreover, from (6), (11) and (21) we find

$$\theta_\psi = -(q_{pp} - q\theta_p^2)\,dp/ds. \tag{23}$$

If we rule out temporarily those flows for which $\theta_\psi \equiv 0$, we may use (23) to eliminate dp/ds from (22) to obtain

$$x_p = -\{(q_{pp} - q\theta_p^2)/\theta_\psi\}\cos\theta, \qquad y_p = -\{(q_{pp} - q\theta_p^2)/\theta_\psi\}\sin\theta. \tag{24}$$

On the other hand we find from (6) and (21) that

$$x_\psi = (q\sin\theta)_p, \qquad y_\psi = -(q\cos\theta)_p. \tag{25}$$

When x and y are eliminated from (24), (25) by partial differentiation, it follows that θ must be a solution of the quasi-linear partial differential equation

$$q(\{q_{pp} - q\theta_p^2\}/\theta_\psi)_\psi + (q^2\theta_p)_p = 0. \tag{26}$$

Conversely, if θ is a solution of (26) for a prescribed Bernoulli function $q = q(p,\psi)$, it is possible to construct a solution (14) of the determinate system (6) in which the density ρ is given by (19). Guided by (21), (24), (25) one tentatively writes down

$$\begin{aligned} x &= \int\{-[(q_{pp} - q\theta_p^2)/\theta_\psi]\cos\theta\,dp + (q\sin\theta)_p d\psi\}, & u &= q\cos\theta,\\ y &= \int\{-[(q_{pp} - q\theta_p^2)/\theta_\psi]\sin\theta\,dp - (q\cos\theta)_p d\psi\}, & v &= q\sin\theta, \end{aligned} \tag{27}$$

and then verifies that this is indeed a solution of (6) by direct substitution.

We return to the case $\theta_\psi \equiv 0$ excluded above. For such flows $\theta = \theta(p)$, i. e. the velocity vectors are all parallel along an isobar. These flows will be examined in detail in Section **5**. At present we are only interested in seeing how (26), (27) must be modified to account for them.

From (23) it follows that (26) is replaced by

$$q_{pp} - q\theta_p^2 = 0. \tag{28}$$

In place of (22) we write

$$x_p = \lambda\cos\theta, \qquad y_p = \lambda\sin\theta, \qquad \lambda = \lambda(p,\psi), \tag{29}$$

and eliminating x, y from (25), (29) by partial differentiation, one finds

$$\lambda = \int_{\psi_0}^{\psi} (q^2\theta_p)_p/q\,d\psi + B(p), \qquad \psi_0 = \text{const.}, \tag{30}$$

where $B(p)$ denotes an arbitrary function. Consequently, in place of (27), we have

$$
(31)\qquad \begin{aligned} x &= \int \{\lambda \cos\theta \, dp + (q \sin\theta)_p d\psi\}, & u &= q\cos\theta; \\ y &= \int \{\lambda \sin\theta \, dp - (q\cos\theta)_p d\psi\}, & v &= q\sin\theta. \end{aligned}
$$

Conversely, if $\theta = \theta(p)$ satisfies (28) for a given Bernoulli function $q = q(p, \psi)$, and λ is defined by (30), the line integrals in (31) define x, y as functions of p, ψ. By direct substitution we verify that (31) is a solution (14) of (6) with the density function ρ given by (19).

We summarize these results in the next theorem.

THEOREM 4. *If the Bernoulli function $q(p, \psi)$ of the flow is given, the stream lines and their hodographs are presented parametrically in terms of the pressure p by* (27), *where θ is a solution of the quasi-linear equation* (26), *provided the direction of flow is not constant along the isobars, in which event* (27) *must be replaced by* (31), *and* (26) *by* (28).

When the partial differentiations are carried out in (26) this equation becomes

$$
(32)\qquad (q_{pp} - q\theta_p{}^2)\theta_{\psi\psi} + 2q\theta_p\theta_\psi\theta_{p\psi} - q\theta_\psi{}^2\theta_{pp} - (q_{pp\psi} - q_\psi\theta_p{}^2 + 2q_p\theta_p\theta_\psi)\vartheta_\psi = 0.
$$

This brings out the quasi-linear character of (26).

5. The case $\boldsymbol{\theta = \theta(p)}$. Equation (28) may be made to play a dual rôle. Depending on whether one elects to preassign the Bernoulli function q or the function θ, it becomes an equation for θ, or an equation for q.

Let us take the equation in the first rôle under the hypothesis that the flow is irrotational. According to Theorem 2 we prescribe

$$
(33)\qquad q = q(p),
$$

and, on solving (28) for θ, obtain

$$
(34)\qquad \theta = \theta(p) = \int_{p_0}^{p} (q_{pp}/q)^{\frac{1}{2}} dp + \theta_0, \qquad \theta_0 = \text{const.}
$$

For θ to be real we admit only values of p for which $q_{pp} \geqq 0$ and therefore, from Theorem 3, exclude subsonic flows.

Substituting from (33), (34) into (31), the second column yields the parametric equations of a one-parameter family of curves (with parameter θ_0) in the hodograph plane, generated by rotating any member h about the origin. A member h of this family is the hodograph of a flow in the physical plane, i. e., the flow in the physical plane has a one-dimensional hodograph.

As is the case for all irrotational flows, it follows from (33) that the isobars in the hodograph plane are circles concentric with the origin.

To construct the stream lines in the physical plane, we carry out the integration of the line integrals in the first column of (31) along the broken line P_0AP indicated below.

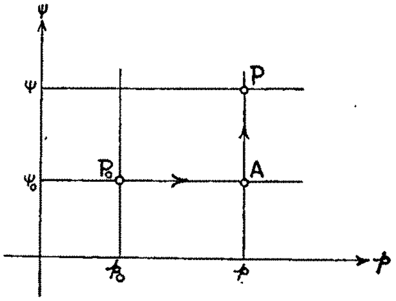


Figure 2.

One finds

(35) $$x = x_0(p) + (\psi - \psi_0)v', \qquad y = y_0(p) - (\psi - \psi_0)u', \quad ' = d/dp,$$

where

(36) $$x_0(p) = \int_{p_0}^{p} B\cos\theta\, dp, \qquad y_0(p) = \int_{p_0}^{p} B\sin\theta\, dp,$$

in which $\theta = \theta(p)$ is defined in (34), and B is the arbitrary function of p introduced in (30).

Setting $\psi = \psi_0$ in (35), one obtains an "initial stream line" S_0

(37) $$x = x_0(p), \qquad y = y_0(p),$$

containing the origin of the physical plane as a point at which $p = p_0$, $\theta = \theta_0$. Once S_0 has been drawn, the other stream lines may be constructed as explained below.

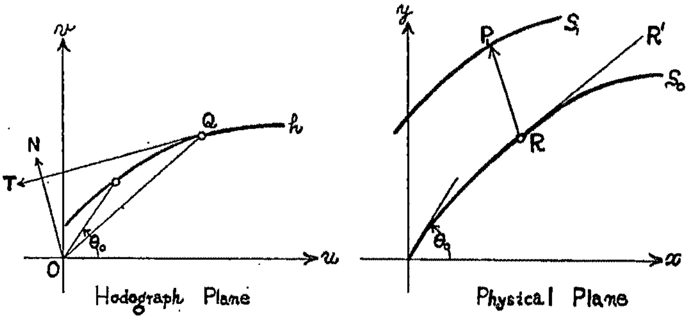


Figure 3.

Through a point R of S_0 draw the tangent RR' and from the origin O of the hodograph plane draw the radius vector $\overrightarrow{OQ}$ parallel to RR' terminating at Q on h. $\overrightarrow{OQ}$ represents the velocity vector at R. At Q draw the tangent vector $\overrightarrow{QT}$ to h with components (u', v') and from O draw the normal vector $\overrightarrow{ON}$ with components $(v', -u')$. From R draw $\overrightarrow{RP_1}$ parallel and equal in length to $\overrightarrow{ON}$. It follows from (35) that the locus S_1 of P_1 as R traverses S_0 is the stream line for which $\psi - \psi_0 = 1$, and that the isobars are the straight line containing the vectors $\overrightarrow{RP_1}$. Along an isobar the velocity vectors are all parallel and equal in length, and the remaining stream lines may be viewed as generalized homothetic transformations of S_1 in which the radii of similitude are the isobars and the center of similitude R moves along S_0.

Due to the arbitrary function $B(p)$ in (36), we may elect any hodograph curve h and still retain wide latitude in the choice of the initial stream line S_0. Fixing p_0, θ_0 in (34), it is possible to make S_0 coincide with any curve C passing through the origin of the physical plane and making an angle θ_0 with the x-axis, provided C has a finite radius of curvature R, e. g., is not a straight line. To see this we take

$$C:\quad x = \int_{\theta_0}^{\theta} R\cos\theta\, d\theta,\qquad y = \int_{\theta_0}^{\theta} R\sin\theta\, d\theta,\qquad R = R(\theta),$$

as parametric equation for C, and using (34) to introduce p as parameter on C; it follows from (36) that S_0 will coincide with C, provided we take $B = R(\theta(p))\theta'(p)$.

If we designate flows for which $\theta = \theta(p)$ as *isoclinic* and flows which have a one-dimensional hodograph as *Prandtl-Meyer flows*, we have seen that any irrotational, isoclinic flow is necessarily a Prandtl-Meyer flow in which the isobars are straight lines along which the velocity vectors are parallel and equal in length. Such flows have been studied in detail by various authors[6] for the special case of a polytropic gas.

There exist, however, rotational, isoclinic flows which are not Prandtl-Meyer flows. To study them, we return to (28) and examine it in its second rôle. From this standpoint we set

$$(38)\qquad\qquad \theta = \theta(p),$$

where $\theta(p)$ is an arbitrarily chosen function, and regard (28) momentarily as an ordinary differential equation for q. If q_1, q_2 denote two linearly independent solutions of (28), the general solution is

---

[6] See, for example, R. Courant and K. O. Friedrichs, *Supersonic flow and shock waves*, pp. 259-293. Interscience Publishers, New York (1948).

$$q = a_1 q_1 + a_2 q_2, \tag{39}$$

where a_1, a_2 denote arbitrary constants, if (28) is regarded as an ordinary differential equation, but denote arbitrary functions of ψ when (28) is interpreted as a partial differential equation.

If we specify the arbitrary functions $a_1(\psi)$, $a_2(\psi)$ in (39), we obtain a definite Bernoulli function $q(p, \psi)$, corresponding to which the function $\theta(p)$, assigned in (38), yields a solution of (28). If we substitute these functions for q, θ in (30), (31), the second column of (31) yields a one-parameter family of hodograph curves

$$h: \quad u = a_1(\psi) u_1 + a_2(\psi) u_2, \qquad v = a_1(\psi) v_1 + a_2(\psi) v_2,$$

where

$$h_1: \quad u_1 = q_1 \cos\theta, \quad v_1 = q_1 \sin\theta; \qquad h_2: \quad u_2 = q_2 \cos\theta, \quad v_2 = q_2 \sin\theta,$$

represent two "base curves" in the hodograph plane which serve to determine the one parameter family of hodograph curves as indicated in Figure 4.

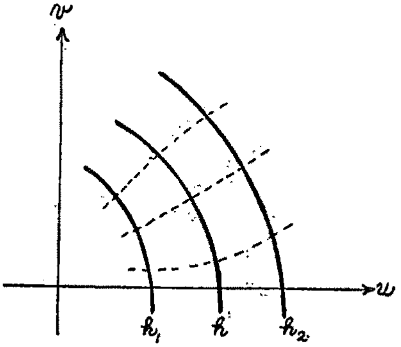


Figure 4.

It follows from (39) and Theorem 2 that our isoclinic flow will be irrotational if, and only if, a_1, a_2 are constant, i. e., if, and only if, h consists of a single curve. Consequently an isoclinic flow will be irrotational if, and only if, it is a Prandtl-Meyer flow.

According to the Munk-Prim Substitution Principle,[7] if u, v denote the velocity components for a given flow, the flow with velocity components λu, λv has the same stream lines and isobars as the given flow, provided $\lambda = \lambda(\psi)$. If two such flows are termed *equivalent*, it follows that all flows equivalent to a given isoclinic flow are isoclinic. Moreover an isoclinic flow will be equivalent to an irrotational flow if, and only if, the functions a_1, a_2 in (39) are linearly dependent. Consequently the necessary and sufficient condition for an isoclinic flow to be equivalent to an irrotational, Prandtl-Meyer flow is that a_1, a_2 be linearly dependent.

---

[7] M. M. Munk and R. C. Prim, *On the multiplicity of steady gas flows having the same stream line pattern.* NOLM 9271, Naval Ordnance Laboratory (1947).

The isobars in the hodograph plane are represented by the dotted lines in Figure 4. To plot them p is fixed in h, and ψ allowed to vary.

The family

$$H: \quad U = A_1(\psi)u_1 + A_2(\psi)u_2, \qquad V = A_1(\psi)v_1 + A_2(\psi)v_2,$$

of curves in the hodograph plane, derived from h by setting

$$(40) \qquad A_1 = \int_{\psi_0}^{\psi} a_1(\psi)\,d\psi, \qquad A_2 = \int_{\psi_0}^{\psi} a_2(\psi)\,d\psi, \qquad \psi_0 = \text{const.}$$

enables one to draw the stream lines in the physical plane. Carrying out the integration of the line integrals in the first column of (31) along the broken line P_0AP in Figure 2, one obtains

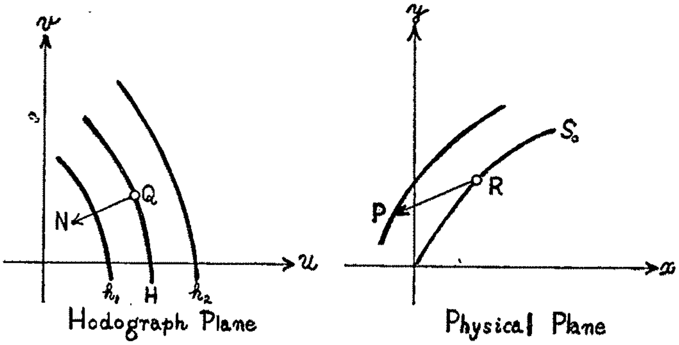


Figure 5.

$$(41) \qquad x = x_0(p) + V_p, \qquad y = y_0(p) - U_p.$$

Here

$$(42) \qquad x_0(p) = \int_{p_0}^{p} B\cos\theta\,dp, \qquad y_0(p) = \int_{p_0}^{p} B\sin\theta\,dp,$$

where θ is the function of p introduced in (38) and B is the arbitrary function of p in (30). From (40) both U_p, V_p vanish for $\psi = \psi_0$ so that

$$S_0: \quad x = x_0(p), \qquad y = y_0(p),$$

is the "initial stream line" $\psi = \psi_0$. To construct the other stream lines we refer to Figure 5.

Choosing a value of ψ, one draws the corresponding member H of the above family and selects a point R on S_0. Corresponding to R, we select a point Q on H so that p has the same value at both points, and erect the normal $\overrightarrow{QN}$ with components $(V_p, -U_p)$ to H at Q. It follows from (41) that the tip P of the vector $\overrightarrow{RP}$ drawn from R, equal in length and parallel to $\overrightarrow{QN}$, traces out the stream line for the selected value of ψ as R traverses S_0.

Because of the arbitrary function $B(p)$ in (42), once the Bernoulli function (39) has been fixed, it is still possible to take the initial stream line S_0 to be any curve through the origin of the physical plane, provided its angle of inclination there equals $\theta(p_0)$ and its radius of curvature is always finite.

An isoclinic flow is never subsonic, since $q_{pp} \geqq 0$ holds from (28).

Taking the Bernoulli function (39) for isoclinic flows, it follows from (19) that the density function necessarily has the form

$$\rho = -(a_1 q_1 + a_2 q_2)^{-1}(a_1 q'_1 + a_2 q'_2)^{-1},$$

where a_1, a_2 denote arbitrary functions of ψ and q_1, q_2 are linearly independent solutions of $q'' - \theta'^2 q = 0$.

Conversely, if we start with such a density function, (6) will permit solutions (14) for which $\theta = \theta(p)$, i. e., isoclinic flows are possible. To prove this, one substitutes for ρ in (15) and disposes the arbitrary functions $p_0(\psi)$, $q_0(\psi)$ so that the Bernoulli function is given by (39). Substituting for q, θ in (30), (31) one obtains solutions (14) of (6) of the desired type.

Let us assume that the entropy distribution function $S = S(\psi)$ can be inverted, $\psi = \psi(S)$ so that a_1, a_2 can be regarded as functions of S. If we designate these functions by σ_1^{-1}, $\sigma_2\sigma_1^{-1}$ we can state the following theorem.

THEOREM 5. *For a fluid to permit isoclinic flows for which* $\theta = \theta(p)$ *under the hypotheses of Theorem* 1, *it is necessary and sufficient that the equation of state has the form*

$$\rho = -\sigma_1^2(q_1 + \sigma_2 q_2)^{-1}(q'_1 + \sigma_2 q'_2)^{-1}, \qquad 1 = d/dp,$$

where q_1, q_2 *denote linearly independent solutions of* $q'' - \theta'^2 q = 0$, *and* σ_1, σ_2 *are arbitrary functions of the entropy* S, *subject to the conditions* $q > 0$, $\rho > 0$.

As an application of this theorem, consider the simplest case possible, $\theta = \text{const.}$ The stream lines are parallel straight lines and we may take $q_1 = -p$, $q_2 = 1$. Consequently the Bernoulli function must be $q = a_2 - a_1 p$, a linear function of p, and the equation of state must have the form $\rho = \sigma_1^2(\sigma_2 - p)^{-1}$. Naturally other forms for the equation of state are possible if the gas flows on parallel straight lines under constant pressure; but if we presuppose the conditions of Theorem 1 (which require a change of pressure along a stream line) only gases having an equation of state of the above form are admissible.

The equation of state (12) for a polytropic gas is *separable,* i. e., it has the form $\rho = \Sigma(S)\Pi(p)$. For the equation of state in Theorem 5 to be separable it is clearly sufficient that $\sigma_2 = \text{const}$. Under this assumption the functions a_1, a_2 in (39) are linearly dependent and the flow is equivalent to an irrotational Prandtl-Meyer flow.

Conversely, we shall show that this is the only case in which the equation of state is separable. In the proof we make use of the lemma below. This lemma is a generalization of a lemma frequently used in solving partial differential equations by separation of variables which states that, for $n = 1$, the identity $\sum_{i=1}^{n} x_i(u)y_i(v) \equiv 1$ implies that $x_1(u)$ and $y_1(v)$ are reciprocal constants. For simplicity, and because it is adequate for our purpose, we restrict attention to the case $n = 3$, although no doubt the lemma can be extended to all positive integral values of n.

LEMMA 1. *In order that two curves*

$$C: \quad x_1 = x_1(u), \qquad x_2 = x_2(u), \qquad x_3 = x_3(u),$$
$$D: \quad y_1 = y_1(v), \qquad y_2 = y_2(v), \qquad y_3 = y_3(v),$$

satisfy the relation $x_1y_1 + x_2y_2 + x_3y_3 = 1$ *identically in* u, v, *it is necessary and sufficient that one curve be a point and the other lie on the polar plane of this point with respect to the unit sphere, or that both curves lie on straight lines which are transforms of each other under polar reciprocation, i. e., either*

(i) *C is a point* (a_1, a_2, a_3) *and D a curve on the plane* $a_1y_1 + a_2y_2 + a_3y_3 = 1$, *or*
D is a point (b_1, b_2, b_3) *and C a curve on the plane* $b_1x_1 + b_2x_2 + b_3x_3 = 1$,

or the curves C, D are presented parametrically by

$$\text{(ii)} \quad \begin{aligned} C:&\quad x_1 = a_1 + Ul_1,\ x_2 = a_2 + Ul_2,\ x_3 = a_3 + Ul_3, \qquad U = U(u),\\ D:&\quad y_1 = b_1 + Vm_1,\ y_2 = b_2 + Vm_2,\ y_3 = b_3 + Vm_3, \quad V = V(v), \end{aligned}$$

where a_i, b_i, l_i, m_i denote constants satisfying the relations

$$a_1b_1 + a_2b_2 + a_3b_3 = 1, \qquad a_1m_1 + a_2m_2 + a_3m_3 = 0,$$
$$l_1b_1 + l_2b_2 + l_3b_3 = 0, \qquad l_1m_1 + l_2m_2 + l_3m_3 = 0.$$

For the proof of this lemma we employ a geometrical argument. To a fixed point $B(b_1, b_2, b_3)$ of D there corresponds the polar plane $b_1x_1 + b_2x_2 + b_3x_3 = 1$ of B with respect to the unit sphere. Since C, D satisfy the relation $x_1y_1 + x_2y_2 + x_3y_3 = 1$ identically in u, v, the curve C must lie in this polar plane. If D consists of a single point B, the identity is fulfilled if, and only if, C is a curve in the polar plane of B and the case comes under alternative (i). If D is not a single point the polar planes corresponding to the points of D constitute a one-parameter family of planes. For the identity to be fulfilled, it is necessary and sufficient that C be in the set S of points common to these planes. For the set S to be not vacuous, it is necessary and sufficient that D be a straight line or a plane curve. In the first case the polar planes are co-axial to yield alternative (ii), and in the latter case they have a single point (a_1, a_2, a_3) in common, as in alternative (i).

Returning to the proof of the necessity of the condition $\sigma_2 = \text{const.}$, in order that the equation of state be separable, it is clear that the product $(q_1 + \sigma_2q_2)(q'_1 + \sigma_2q'_2)$ must be separable, consequently the identity in Lemma 1 must be satisfied by the curves

$$C: \quad x_1 = -(q_1q_2)'/q_1q'_1 = x_1(p), \quad x_2 = -q_2q_2'/q_1q'_1 = x_2(p), \quad x_3 = x_3(p),$$

$$D: \quad y_1 = \sigma_2 = y_1(S), \qquad y_2 = \sigma_2^2 = y_2(S), \qquad y_3 = y_3(S).$$

If C were a point (a_1, a_2, a_3), it would follow that

$$q_1q_2 = -\tfrac{1}{2}a_1q_1^2 + c_1, \qquad q_2^2 = -a_2q_1^2 + c_2, \qquad c_1 = \text{const.}, \; c_2 = \text{const.},$$

and hence that q_1, q_2 are both constant or else linearly dependent, neither of which is possible, since q_1, q_2 designate linearly independent solutions of $q'' - \theta'^2q = 0$. Thus C cannot be a point. The projection of D on the plane $y_3 = 0$ is the parabola $y_1^2 = y_2$, so D cannot contain a straight line segment. Consequently D is a point, and σ_2 is a constant as stated.

These results lead to the following theorem.

THEOREM 6. *The rotational, isoclinic flows of a gas with a separable equation of state subject to the conditions in Theorem* 1 *are equivalent to irrotational, Prandtl-Meyer flows under the Munk-Prim Substitution Principle.*

6. The case $\theta = \theta(\psi)$. Equation (26), like (28), plays two rôles, depending on whether q, or θ, is taken to be an assigned function of p, ψ. Taking (26) in its second rôle, we prescribe $\theta = \theta(\psi)$ and take q for unknown function. We suppose $\theta' \neq 0$, so that the stream lines lie on a family of non-parallel straight lines and term such flows *rectilinear* flows. For rectilinear flows (26) reduces to $(q_{pp}/\theta_\psi)_\psi = 0$, the general solution of which

$$(43) \qquad q = \dot{\theta}[P(p) + \Psi_1(\psi)p + \Psi_2(\psi)], \qquad \dot{\theta} \neq 0, \qquad (\cdot = d/d\psi),$$

involves three arbitrary functions $P(p)$, $\Psi_1(\psi)$, $\Psi_2(\psi)$. If we specify these arbitrary functions, we obtain a definite Bernoulli function $q(p,\psi)$ corresponding to which the prescribed function $\theta(\psi)$ yields a solution of (26), and, on substituting for q, θ in (27), we obtain a solution (14) of (6) for which the density function is given by

$$(44) \qquad \rho = -\dot{\theta}^{-2}(P' + \Psi_1)^{-1}(P + \Psi_1 p + \Psi_2)^{-1}.$$

Conversely, if we start with a density function of this type, (6) will permit solutions (14) for which $\theta = \theta(\psi)$, i. e., rectilinear flows are possible. To obtain them, one substitutes from (44) into (15) and adjusts the arbitrary functions $p_0(\psi)$, $q_0(\psi)$ so that the Bernoulli function is given by (43). The desired solution (14) of (6) is then obtained by substituting for q, θ in (27).

Introducing the entropy S as in Section **5**, we obtain the theorem in view of (23).

THEOREM 7. *For a fluid to permit rectilinear flows under the hypotheses of Theorem* 1, *it is necessary and sufficient that the equation of state has the form*

$$(45) \qquad \rho = -\sigma^2(P' + \sigma_1)^{-1}(P + \sigma_1 p + \sigma_2)^{-1}, \qquad ' \doteq d/dp,$$

where P is any non-linear function of the pressure p and σ, σ_1, σ_2 denote arbitrary functions of the entropy S, subject to the conditions $q > 0$, $\rho > 0$.

Further insight into rectilinear flows is given by the following theorem.

THEOREM 8. *Under the hypotheses of Theorem* 1, *the necessary and sufficient conditions for the isobars to be orthogonal trajectories of the stream lines is that the stream lines lie on straight lines.*

If we let K denote the curvature of a stream line this theorem is an immediate consequence of the relation $x_p x_\psi + y_p y_\psi = qK(ds/dp)^2$, derived by substituting for x_p, y_p from (22) and for x_ψ, y_ψ from (6) with the help of (21).

Consequently if the stream lines lie on straight lines L the isobars are the involutes I of the evolute E enveloped by the straight lines.

To confirm and elaborate upon this result, we substitute for q from (43) and put $\theta = \theta(\psi)$ in (27) to obtain

$$x = \int \Psi_1 \sin\theta d\theta - P' \cos\theta, \qquad y = -\int \Psi_1 \cos\theta d\theta - P' \sin\theta.$$

and this, upon integration by parts, yields

$$(46) \quad \begin{aligned} x &= \int (\dot{\Psi}_1/\dot{\theta}) \cos\theta d\theta - (P' + \Psi_1)\cos\theta, \\ y &= \int (\dot{\Psi}_1/\dot{\theta}) \sin\theta d\theta - (P' + \Psi_1)\sin\theta. \end{aligned}$$

If we let R denote the radius of curvature of E and τ be the arc length of E measured from some fixed point P_0 on E, the equations of the involutes of E are

$$(47) \qquad x = \int R \cos\theta d\theta + (a - \tau)\cos\theta, \qquad y = \int R \sin\theta d\theta + (a - \tau)\sin\theta,$$

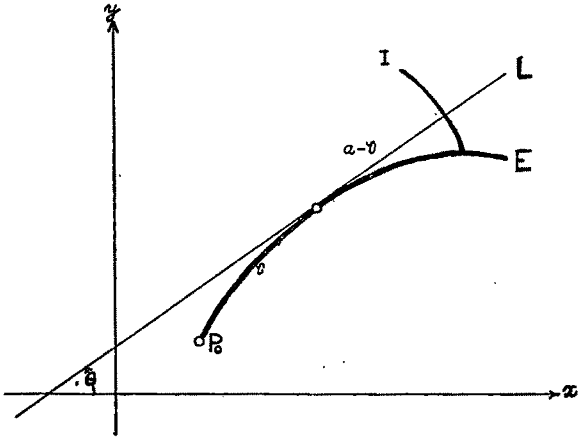


Figure 6.

in which a specific involute is obtained by assigning a definite value to the constant a. Comparing (47) with (46), we set

$$(48) \qquad R = \dot{\Psi}_1/\dot{\theta}, \qquad \tau = \psi_1, \qquad a = -P'.$$

Consequently the stream lines in (46) lie on straight lines enveloping a curve E whose intrinsic equation $R = R(\tau)$ is obtained by eliminating ψ from the first two equations in (48). It follows that E is uniquely determined up to rotations and translations by prescribing the arbitrary function $\Psi_1(\psi)$. The third equation in (48) determines the pressure to be assigned to each involute I in its rôle as an isobar, provided, of course that $P' \neq$ const. From (43), (44) it is clear that neither the lines of constant speed (isovels) nor the lines of constant density (isopycnics) are necessarily the same as the isobars, as is always the case in irrotational flows.

The arbitrary function $\Psi_1(\psi)$ may be chosen so that E is any curve in the physical plane with a continuous radius of curvature, for if $R = R(\tau)$ is the intrinsic equation of E, it is sufficient, from (48), to take any solution of the differential equation $\dot{\Psi}_1 = \theta R(\Psi_1)$ for Ψ_1.

Concurrent straight lines are characterized by $R = 0$, and therefore, from (48), by the condition $\Psi_1 =$ const.

Referring to Figure 6, if we denote the distance $a - \tau$ measured along L from its point of tangency with E to the isobar I by s, it follows from (48), (43) that $s = -q_p/\dot{\theta}$. Consequently if p passes through a value p^* at which $\dot{q}_{pp}$ changes sign, the flow changes from subsonic to supersonic or vice versa and s passes through a maximum or a minimum. This would lead to two different sets of flow quantities for the same point of L, which is impossible. It follows that p cannot be allowed to pass through any such value p^* and consequently that a stream line need not occupy all of the straight line L upon which it lies, at least insofar as our analysis goes.[8]

Summing up these results we have the theorem.

THEOREM 9. *The stream lines in the rectilinear flow of a gas having the density function* (44) *with preassigned functions* $\dot{\theta}$, Ψ_2 *may be made to lie on the straight lines enveloping any given curve* E *in the physical plane with a continuous radius of curvature by properly choosing the arbitrary function* Ψ_1. *In particular, the straight lines will be concurrent if, and only if* $\Psi_1 = const.$

As in Section **5**, the equation of state (45) will be separable if, and only if, the product $(P' + \sigma_1)(P + \sigma_1 p + \sigma_2)$ is separable.

LEMMA 2. *The product* $(P' + \sigma_1)(P + \sigma_1 p + \sigma_2)$ *is separable, if, and only if, one of the following alternatives holds.*

$$\text{(i)}\quad \sigma_1 = c_1,\quad P = -c_1 p + c_2; \qquad \text{(ii)}\quad \sigma_1 = c_1,\quad \sigma_2 = c_2;$$
$$\text{(iii)}\quad P = c_1 p + c_2,\quad c_2 + \sigma_2 = k(c_1 + \sigma_1),$$

where c_1, c_2, k *denote constants.*

It is a simple matter to verify that any one of these alternatives is sufficient.

To prove that the three alternatives exhaust the possibilities, we exclude (i) and show that either (ii) or (iii) must hold.

Under the hypothesis $\sigma_1 = c_1$ the product $(P' + c_1)(P + c_1 p + \sigma_2)$ must be separable and, since (i) is excluded, the factor $P + c_1 p + \sigma_2$ must be separable, which can only be the case if $\sigma_2 = \text{const.}$, i. e., if (ii) holds.

If, on the other hand, $\dot{\sigma}_1 \neq 0$, we differentiate the identity

$$(48)\qquad (P' + \sigma_1)(P + \sigma_1 p + \sigma_2) = PP' + (pP')\sigma_1 + P'\sigma_2 + p\sigma_1^2 + \sigma_1\sigma_2,$$

---

[8] S. Bergman, *The hodograph method in the theory of compressible fluid.* Brown University Publication, Providence (1941), in which the flow of a polytropic gas for a compressible source or sink is investigated and the analysis is restricted to the exterior of a "limit circle."

partially with respect to p and S, and, since the result must be separable, we may write $(pP)''\dot{\sigma}_1 + P''\dot{\sigma}_2 + 2\sigma_1\dot{\sigma}_1 = \Pi\Sigma$, $\Pi = \Pi(p)$, $\Sigma = \Sigma(S)$, which, on dividing by $2\sigma_1\dot{\sigma}_1$ may be replaced by $-(pP)''(2\sigma_1)^{-1} - P''\dot{\sigma}_2(2\sigma_1\dot{\sigma}_1)^{-1} + \Pi\Sigma_1 = 1$, $\Sigma_1 = (2\sigma_1\dot{\sigma}_1)^{-1}\Sigma$, to which we apply Lemma 1. Since $\sigma_1 \neq$ const., either one of

$$\text{(a)} \quad -(pP)'' = a_1, \qquad -P'' = a_2, \qquad \Pi = a_3,$$

$$\text{(b)} \quad (2\sigma_1)^{-1} = b_1 + Vm_1, \quad \dot{\sigma}_2(2\sigma_1\dot{\sigma}_1)^{-1} = b_2 + Vm_2, \quad \Sigma_1 = b_3 + Vm_3,$$

must hold. Under (a) the first two equations imply $a_2 = 0$. Consequently $P = c_1 p + c_2$ and therefore

$$(P' + \sigma_1)(P + \sigma_1 p + \sigma_2) = (c_1 + \sigma_1)[(c_1 + \sigma_1)p + c_2 + \sigma_2].$$

For this to be separable it is necessary that $c_2 + \sigma_2 = k(c_1 + \sigma_1)$, and we arrive at alternative (iii).

Taking up (b), we eliminate V from the first two equations and integrate to obtain

$$(49) \qquad m_1\sigma_2 = \mu\sigma_1^2 + m_2\sigma_1 + c_1, \qquad \mu = m_1 b_2 - m_2 b_1, \qquad c_1 = \text{const.}$$

where $m_1 \neq 0$, since if $m_1 = 0$, the first equation in (b) implies $\sigma_1 =$ const. If we use (49) to eliminate σ_2 from (48), the identity becomes

$$(50) \quad m_1(P' + \sigma_1)(P + \sigma_1 p + \sigma_2) = (c_1 + m_1 P)P' + \{c_1 + [(m_1 p + m_2)P]'\}\sigma_1 + (m_1 p + m_2 + \mu P')\sigma_1^2 + \mu\sigma_1^3.$$

Differentiating partially with respect to S, we obtain a necessary condition for separability which, provided $\mu \neq 0$, may be identified with the identity in Lemma 1 for the curves

$$C\colon \quad x_1 = -c_1 - [(m_1 p + m_2)P]', \qquad x_2 = -(m_1 p + m_2 + \mu P'), \qquad x_3 = x_3(p),$$

$$D\colon \quad y_1 = (3\mu\sigma_1^2)^{-1}, \qquad y_2 = 2(3\mu\sigma_1)^{-1}, \qquad y_3 = y_3(S).$$

Clearly D cannot be a point or straight line segment, so C must be a point. Now x_1, x_2 can both be constant only if $m_1 = 0$, and since actually $m_1 \neq 0$, the assumption $\mu \neq 0$ leads to a contradiction and must be rejected.

For $\mu = 0$, the condition for separability of the product in (50) assumes the form of the identity in Lemma 1 for the curves.

$$C\colon \quad x_1 = -(c_1 + m_1 P)P'/(m_1 p + m_2),$$
$$x_2 = -\{c_1 + [(m_1 p + m_2)P]'\}/(m_1 p + m_2), \quad x_3 = x_3(p),$$

$$D\colon \quad y_1 = \sigma_1^{-2} \qquad y_2 = \sigma_1^{-1}, \qquad y_3 = y_3(S).$$

As above D cannot be a point or straight line segment, so C must be a point. This time, however, for x_1, x_2 to reduce to constants it is necessary that P be a linear function of p and proceeding as in (a), we are led to alternative (iii).

Lemma 2 is used to prove the following theorem.

THEOREM 10. *Under the hypotheses of Theorem* 1 *the stream lines in the rectilinear flow of a gas with a separable equation of state must lie on concurrent straight lines and the flow upon them is equivalent to an irrotational flow under the Munk-Prim Substitution Principle.*

For a rectilinear flow with the Bernoulli function (43) it is easy to verify that (23) reduces to $dp/ds = -1/P''$, and, if the pressure gradient is to remain finite for a gas with a separable equation of state, alternatives (i), (iii) in Lemma 2 are ruled out. Under alternative (ii) both σ_1, σ_2 are constant in (45), the straight lines are concurrent by Theorem 9, and the flow upon them equivalent, from (43), to an irrotational flow under the Munk-Prim Substitution Principle.

UNIVERSITY OF MARYLAND,
COLLEGE PARK.

ON THE ASYMPTOTIC SHAPE OF THE CAVITY BEHIND AN AXIALLY SYMMETRIC NOSE MOVING THROUGH AN IDEAL FLUID.*

By Francis Scheid.

Introduction. This paper continues and in a sense completes that of Levinson [1] in which it is shown that the application of Green's theorem to the problem described in the title leads to a singular, non-linear integro-differential equation which can be written in the form [2]

$$(3.10) \qquad \phi(x, f(x)) = I_1 + I_2 + I_3 + J_1 + J_2 + J_3 + \text{const.} + O(1/x),$$

where x is distance from the tip of the nose measured along the axis of symmetry, $r = f(x)$ is the equation of the free boundary, ϕ is a harmonic function such that $V + \phi_x$ and ϕ_r are the x and r components of the fluid velocity, V being the velocity at infinity, and the remaining symbols denote surface integrals whose exact form will appear as we proceed.

The problem of cavity shapes in two dimensions has received considerable attention, the principle technique being that of conformal mapping. However, in the three-dimensional case no general method is known. For the problem at hand Levinson's result is

$$(1.4) \qquad f(x) = (cx^{1/2}/\log^{1/4} x)\,[1 - (\log\log x)/(8\log x) + O(1/\log x)],$$

where x is large and c constant. In obtaining this result certain assumptions are made involving the smoothness of $f(x)$ leading to an asymptotic appraisal of the integrals in (3.10) which then gives a weighted mean of $f(x)$. (1.4) is then a consequence of what amounts to a Tauberian theorem.

Here we make the additional assumption that the unknown quantity in (1.4) has a derivative which is $O(1/x\log^2 x)$ and find by differentiation

$$(1) \quad f'(x) = (f(x)/2x)\,[1 - 1/(2\log x) + (\log\log x)/(4\log^2 x) + O(1/\log^2 x)].$$

In Part I which follows, the integrals in (3.10) are reduced to a form which makes their reappraisal a matter of substitution and computation. Details are provided so that in Part II we may extend (1.4) to

---

* Received June 9, 1949.

$$f(x) = (cx^{1/2}/\log^{1/4} x)[1 - (\log\log x)/(8\log x) + (1/4 + \tfrac{1}{2}\log(\tfrac{1}{2}c))/\log x + (5/128)(\log\log x/\log x)^2 - (7/32 + (5/16)\log\tfrac{1}{2}c)(\log\log x/\log^2 x) + O(1/\log^2 x)].$$

The nature of the computations necessary to continue this series makes it apparent that the shape of the nose, except for the general assumptions already made, has no effect upon the asymptotic shape of the cavity wall.

Other equations obtained by Levinson which we shall use are:

$$\phi_n = (f'(x))(1 + f'^2(x))^{-\frac{1}{2}}; \tag{2.17}$$

$$f(x+y) < 2f(x), \qquad \text{if } |y| \leq \tfrac{1}{2}x; \tag{3.5}$$

$$\int\int_{w_2} \cos\psi/R^2\, dS = -2\pi + O(f^2(x)/x^2); \tag{3.15}$$

$$\phi(x, f(x)) = C_2 + \tfrac{1}{2}B(x) + O(x^{-\frac{1}{2}}); \tag{6.5}$$

$$J_1 = O(x^{-1+\epsilon}); \tag{6.6}$$

$$J_3 = O(x^{-1+\epsilon}); \tag{6.7}$$

$$B(x) = \int_a^x f'^2(t)\,dt = \text{const.} + E_1 + (1/4)\int_a^x f^2(t)/t^2\,dt, \quad \text{where } E_1 = -\int_x^\infty [f'^2(t) - 1/4\, f^2(t)/t^2]\,dt; \tag{6.15}$$

$$\int_a^x f(t)f'(t)/(t+1)\,dt = \text{const.} + F_1 + \tfrac{1}{2}\int_a^x f^2(t)/t^2\,dt, \quad \text{where } F_1 = -\int_x^\infty [f(t)f'(t)/(t+1) - \tfrac{1}{2}f^2(t)/t^2]\,dt; \tag{6.17}$$

$$R(y) = \int_{\log a}^y H(n)/n^{\frac{1}{2}}\,dn - 2y^{\frac{1}{2}}H(y); \tag{6.19}$$

$$H(y) = c^2 - \tfrac{1}{2}R(y)/y^{\frac{1}{2}} + (1/4)\int_y^\infty R(n)/n^{3/2}\,dn. \tag{6.20}$$

More recently M. I. Gurevich has obtained the same result as Levinson, but in an entirely different manner. Gurevich's paper appears in *Prikladnaia Matematika y Mekhanika*, vol. 11, No. 1, and is titled "Flow past an Axi-symmetric Semi-Body of Finite Drag."

Part I. Reappraisal of the Integrals.

We consider first $I_1 = -\frac{1}{2\pi}\int\int_{w_1}(1/R - 1/(\xi+1))\phi_n dS$. The second term gives, using (2.17), (1), and (1.4),

$$\left(\int_a^x - \int_{3x/4}^x\right)[f(\xi)f'(\xi)/(\xi+1)]\,d\xi + \left(\int_a^\infty - \int_{3x/4}^\infty\right)[f(\xi)O(f'^3(\xi))/(\xi+1)]\,d\xi,$$

where the last term is of the form const. $+\, O(1/x)$. The first term gives

$$-\tfrac{1}{2\pi}\iint_{w_1} f'(\xi)/R\,dS - \tfrac{1}{2\pi}\iint_{w_1} O(f'^3(\xi))/R\,dS = Q_1 + Q_2.$$

Now $|Q_2| \leq 4/x \int_a^x f(\xi)O(f'^3(\xi))\,d\xi = O(\operatorname{Log} x/x)$.

For Q_1 we have $1/R = [(x-\xi)^2 + r^2 + \rho^2 - 2r\rho\cos(\theta-\lambda)]^{-\frac{1}{2}}$

$$= [1 + O(f^2(x)/x^2)]^{-\frac{1}{2}}/(x-\xi) = 1/(x-\xi) + O(f^2(x)/x^3)$$

and so

$$Q_1 \doteq -\tfrac{1}{2\pi}\iint_{w_1} f'(\xi)/(x-\xi)\,dS + \iint_{w_1} O(f^2(x)/x^3)f'(\xi)\,dS$$

$= Q_3 + Q_4$. But $|Q_4| \leq 2\pi f^2(x)/x^3.\ \text{const.} \int_a^{3x/4} f(\xi)f'(\xi)\,d\xi$

$$< \text{const.}\ f^3(x)/x^3 \int_a^x f'(\xi)\,d\xi = O(1/x).$$

Next we have

$$Q_3 = -\int_a^{3x/4} f(\xi)f'(\xi)/(x-\xi)\,d\xi = \left(-\int_a^b - \int_b^{3x/4}\right)[f(\xi)f'(\xi)/(x-\xi)]\,d\xi$$

$= Q_5 + Q_6$, where $b = x/\log^n x$. Using (1. 4) we find

$$|Q_5| \leq [1/(x-b)]f^2(b) \leq \text{const.}\ b/(\log^{\frac{1}{2}} b)(x-b) = O(\log^{-n-\frac{1}{2}} x).$$

We choose $n \geq 1$. Once again using (1) we have

$$Q_6 = -\int_b^{3x/4} [1 + O(1/\log x)]f^2(\xi)/2\xi(x-\xi)\,d\xi.$$

Since $\int_b^{3x/4} d\xi/(x-\xi)$ is bounded for large x, the error term contributes $O(\log^{-3/2} x)$. Thus

$$Q_6 = -\int_b^{3x/4} f^2(x)/2x(x-\xi)\,d\xi - \int_b^{3x/4} [f^2(\xi)/2\xi - f^2(x)/2x]/(x-\xi)\,d\xi + O(\log^{-3/2} x).$$

$$= -f^2(x)(\log 4)/2x - \int_b^{3x/4} + \int_x^{\xi} [f(t)f'(t)/t - f^2(t)/2t^2]\,dt\,d\xi/(x-\xi) + O(\log^{-3/2} x)$$

$$= -f^2(x)(\log 4)/2x + O(\log^{-3/2} x)$$

$$-\left(\int_x^{3x/4}\int_b^{3x/4} + \int_{3x/4}^{b}\int_b^{t}\right)[f(t)f'(t)/t - f^2(t)/2t^2]\,d\xi/(x-\xi)\,dt$$

$$= -f^2(x)(\log 4)/2x - Q_7 - Q_8 + O(\log^{-3/2} x).$$

In Q_7 the integrations can be performed so that

$$Q_7 = \int_x^{3x/4}(\log 4 + O(\log^{-n} x))\,d/dt(f^2(t)/2t)\,dt$$

$$< \text{const. } [f^2(3x/4)/(3x/4) - f^2(x)/x], \text{ and now using } (1.\,4)$$

$$< \text{const. } [(1/(\log^{1/2} 3x/4) - 1/(\log^{1/2} x))$$

$$-((\log\log 3x/4)/4(\log 3x/4)^{3/2} - (\log\log x)/4\log^{3/2} x)$$

$$+ O(\log^{-3/2} x)]$$

$= O(\log^{-3/2} x)$, where the following identities are used:

(A) $\log\log 3x/4 = \log\log x + O(\log^{-1} x)$

(B) $\log^{-1/2} 3x/4 = \log^{-1/2} x + O(\log^{-3/2} x)$

(C) $\log^{-3/2} 3x/4 = \log^{-3/2} x + O(\log^{-5/2} x)$.

In Q_8 carrying out the ξ integration gives

$Q_8 = \int_{3x/4}^{b}[f(t)f'(t)/t - f^2(t)/2t^2]\log[(x-b)/(x-t)]\,dt$, which by (1)

$$= \int_{3x/4}^{b}(f^2(t)/4t^2\log t)[-1 + (\log\log t/(2\log t))$$

$$+ O(1/\log t)]\log[(x-b)/(x-t)]\,dt.$$

Over this range of integration $\log[(x-b)/(x-t)]$ is bounded; moreover, by (1. 4),

$$f^2(t)/(t\log t) < k/\log^{3/2} t < k/\log^{3/2} b = O(\log^{-3/2} x),$$

k being a constant. Accordingly in Q_8, the last two terms in brackets contribute $O(\log^{-3/2} x)$. Noting also that $\log(x-b) = \log x + O(\log^{-n} x)$ we have

$$Q_8 = \int_b^{3x/4}(f^2(t)/4t^2\log t)(\log x/x - t)\,dt + O(\log^{-3/2} x).$$

Now

$$(f^2(t)/t\log t) - (f^2(x)/x\log x)$$

$$< f^2(x\log^{-n} x)/[x\log^{-n} x\log(x\log^{-n} x)] - f^2(x)/x\log x$$

$$< \text{const.}(\log^{-3/2} b - \log^{-3/2} x) + O(\log^{-2} x) = O(\log^{-2} x),$$

where we have used the identity

(D) $(\log x - n \log\log x)^{-3/2} = \log^{-3/2} x + O(\log^{-2} x)$.

So replacing $f^2(t)/(t \log t)$ by $f^2(x)/(x \log x)$ in Q_8 gives an error of

$$O(\log^{-2} x) \int_b^{3x/4} dt/t = O(\log^{-3/2} x),$$

the logarithm in the integrand being bounded as before. But now

$$Q_8 = (f^2(x)/(4x \log x)) \int_b^{3x/4} \log\,(x/t(x-t))\,dt + O(\log^{-3/2} x),$$

and since $\log x - \log\,(x-t) < kt/x$, the remaining integral is bounded, so that $Q_8 = O(\log^{-3/2} x)$.

In summary then,

$$I_1 = \left(\int_a^x - \int_{3x/4}^x\right) f(\xi)f'(\xi)/(\xi+1)\,d\xi - (\log 4) f^2(x)/2x$$
$$+ \text{const.} + O(\log^{-3/2} x).$$

Next we consider $I_2 = -\frac{1}{2\pi} \int\int_{w_2} [1/R - 1/(\xi+1)]\phi_n\,dS$. Using (2. 17), (1) and (1. 4), the second term becomes

$$\left(\int_{3x/4}^x + \int_x^{5x/4}\right) f(\xi)f'(\xi)/(\xi+1)\,d\xi + O(1/x).$$

The first of these integrals occurs with the opposite sign in I_1 and for the second, using (1) and (1. 4) we find

$$|\,f(x)f'(x) - f(\xi)f'(\xi)\,| < f(x)f'(x) - f(5x/4)f'(5x/4)$$
$$= \tfrac{1}{2}c^2[1/\log^{\frac{1}{2}} x - 1/\log^{\frac{1}{2}}(5x/4) - (\log\log x)/(4 \log^{3/2} x)$$
$$+ (\log\log 5x/4)/(4 \log^{3/2} 5x/4)] + O(\log^{-3/2} x)$$

$= O(\log^{-3/2} x)$ just as before, using identities almost the same as (A), (B), and (C). Accordingly, since $\int_x^{5x/4} d\xi/(\xi+1) = \log 5/4 + O(1/x)$, $f(\xi)f'(\xi)$ may be replaced by $f(x)f'(x)$, introducing an error of $O(\log^{-3/2} x)$. Hence

$$\int_x^{5x/4} f(\xi)f'(\xi)/(\xi+1)\,d\xi = f(x)f'(x) \log 5/4 + O(\log^{-3/2} x),$$

$= (f^2(x)/2x)(\log 5/4) + O(\log^{-3/2} x)$, by (1). The second term in I_2 therefore contributes

$$\int_{3x/4}^{x} f(\xi)f'(\xi)/(\xi+1)\,d\xi + (f^2(x)/2x)(\log 5/4) + O(\log^{-3/2} x).$$

The first term in I_2 is $-\frac{1}{2\pi}\int\int_{w_2}(1/R)\phi_n\,dS$

$$= -\frac{1}{2\pi}\int\int_{w_2} f'(\xi)(1+f'^2(\xi))^{-\frac{1}{2}}[(x-\xi)^2+(r+\rho)^2 - 2r\rho(1+\cos(\theta-\lambda))]^{-\frac{1}{2}}\,dS$$

$$= -\frac{1}{2\pi}\int_{3x/4}^{5x/4} f(\xi)f'(\xi)(1+f'^2(\xi))^{-\frac{1}{2}} \int_0^{\frac{1}{2}\pi} 4[(x-\xi)^2+(r+\rho)^2-4r\rho\cos^2\lambda]^{-\frac{1}{2}}\,d\lambda\,d\xi$$

$$= -\frac{1}{2\pi}\int_{3x/4}^{5x/4} f(\xi)f'(\xi)(1+f'^2(\xi))^{-\frac{1}{2}} \int_0^{\frac{1}{2}\pi} 4(4r\rho)^{-\frac{1}{2}}(\alpha^2+\sin^2\lambda)^{-\frac{1}{2}}\,d\lambda\,d\xi,$$

where we have used (2. 17) and made a change of variable in the inner integral, and where

$$\alpha^2 = ((x-\xi)^2+(r+\rho)^2-4r\rho)/4r\rho = [(x-\xi)^2+(r-\rho)^2]/4r\rho$$
$$= ((x-\xi)^2/4r\rho)(1+(r-\rho)^2/(x-\xi)^2).$$

Now whether $\xi > x$ or $\xi < x$, $(r-\rho)^2/(x-\xi)^2 < f'^2(3x/4)$, so that

$$(2) \qquad \alpha^2 = ((x-\xi)^2/4r\rho)(1+O(f^2(x)/x^2)).$$

It is convenient to split this last integral into the following pieces:

R_1 = part for which $x_1 \leq \xi \leq x_2$;

R_2 = part for which $x_2 \leq \xi \leq x_4$ or $x_3 \leq \xi \leq x_1$;

R_3 = part for which $x_4 \leq \xi \leq 5x/4$ or $3x/4 \leq \xi \leq x_3$,

where x_1, x_2, x_3 and x_4 denote respectively $x-\delta_n$, $x+\delta_n$, $x-x^{3/5}$, and $x+x^{3/5}$.

To treat R_1, we note that $\int_0^{\frac{1}{2}\pi}[(\alpha^2+\sin^2\lambda)^{-\frac{1}{2}}-(\alpha^2+\lambda^2)^{-\frac{1}{2}}]\,d\lambda$

$$= \int_0^{\frac{1}{2}\pi}(\lambda^2-\sin^2\lambda)(\alpha^2+\sin^2\lambda)^{-\frac{1}{2}}(\alpha^2+\lambda^2)^{-\frac{1}{2}}/[(\alpha^2+\sin^2\lambda)^{\frac{1}{2}} + (\alpha^2+\lambda^2)^{\frac{1}{2}}]\,d\lambda$$

$$< \int_0^{\frac{1}{2}\pi}\text{const.}\,\lambda\,d\lambda = \text{const.}$$

Consequently, replacing $(\alpha^2+\sin^2\lambda)^{\frac12}$ by $(\alpha^2+\lambda^2)^{\frac12}$ in R_1 introduces an error of $<$ const. $\delta_n f(x)/x$ (by (1) and (3.5)). This error will be $O(\log^{-n} x)$ if we choose

$$(3)\qquad \delta_n = b/f(x).$$

Hence

$$R_1 = -\tfrac{1}{\pi}\int_{x_1}^{x_2} f(\xi)f'(\xi)[1+f'^2(\xi)]^{-\frac12}(r\rho)^{-\frac12}\int_0^{\frac12\pi}(\alpha^2+\lambda^2)^{-\frac12}d\lambda\, d\xi + O(\log^{-n} x),$$

with δ_n as in (3). But the λ integration can now be performed and

$$R_1 = -\tfrac{1}{\pi}\int_{x_1}^{x_2} f(\xi)f'(\xi)[1+f'^2(\xi)]^{-\frac12}(r\rho)^{-\frac12}\log\{[\tfrac12\pi+(\alpha^2+\pi^2/4)^{\frac12}]/|\,\alpha\,|\}d\xi$$
$$+ O(\log^{-n} x).$$

Now $\log\{[\frac12\pi+(\alpha^2+\pi^2/4)^{\frac12}]/|\,\alpha\,|\} = \log(\pi/|\,\alpha\,|) + O(\alpha^2)$, and observing that by (2), $\alpha^2 = O((x-\xi)^2/r\rho) = O(\delta^2_n/r^2) = O(\log^{1-2n} x)$, we find using (3.5) that the term $O(\alpha^2)$ in R_1 contributes

$$< \text{const.}\ (\log^{1-2n} x)[f(x_2)-f(x_1)]$$
$$< \text{const.}\ (\log^{1-2n} x)\delta_n f'(x_1)\quad \text{(by the mean-value theorem)}$$
$$= O((\log^{1-2n} x)\delta_n f(x)/x_1) = O(\log^{1-3n} x)\ \text{by (1) and (3).}$$

The same argument shows that the $\log\pi$ term contributes $O(\log^{-n} x)$ so that

$$R_1 = \tfrac{1}{\pi}\int_{x_1}^{x_2} f(\xi)f'(\xi)[1+f'^2(\xi)]^{-\frac12}(r\rho)^{-\frac12}(\log|\,\alpha\,|)d\xi + O(\log^{-n} x).$$

By (2),

$$\log|\,\alpha\,| = \tfrac12\log\alpha^2 = \tfrac12\log\{[(x-\xi)^2/4r\rho][1+O(f^2(x)/x^2)]\}$$
$$= \tfrac12\log((x-\xi)^2/4r\rho) + O(1/x).$$

As above the term $O(1/x)$ contributes $O(1/x)$; furthermore by (3.5) and (1.4), $|\log 4r\rho| < \log 8r^2 = O(\log r) = O(\log x)$, so again as before this term contributes $O(\log^{1-n} x)$ and

$$R_1 = \tfrac{1}{\pi}\int_{x_1}^{x_2} f(\xi)f'(\xi)[1+f'^2(\xi)]^{-\frac12}(r\rho)^{-\frac12}(\log|\,x-\xi\,|)d\xi + O(\log^{1-n} x).$$

Now $\int_{x_1}^{x_2}\log|\,x-\xi\,|\,d\xi = 2\int_0^{\delta_n}\log p\,dp = 2(\delta_n\log\delta_n - \delta_n) <$ const. $\delta_n\log x$,

so that $|\,R_1\,| \leq$ const. $\delta_n\log x f(x)/x + O(\log^{1-n} x) = O(\log^{1-n} x)$.

We next turn to

$$R_2 = -\tfrac{1}{2\pi}\left(\int_{x_3}^{x_1}+\int_{x_2}^{x_4}\right) f(\xi)f'(\xi)[1+f'^2(\xi)]^{-\frac12}\int_0^{\frac12\pi}4(4r\rho)^{-\frac12}(\alpha^2+\sin^2\lambda)^{-\frac12}d\lambda d\xi.$$

To simplify this integral we note:

(A) Replacing $[1+f'^2(\xi)]^{-\frac{1}{2}}$ by 1 introduces an error of

$$< \text{const. } x^{3/5}(f(x)/x)^3(f(x)/\delta_n) = O(x^{-\frac{1}{2}}),$$

where (1), (2), (3) and (1.4) have been used.

(B) Replacing $\rho^{\frac{1}{2}}$ by $r^{\frac{1}{2}}$ introduces an error of

$$< \text{const. } x^{3/5}[f^{3/2}(x)/x][f(x)/\delta_n]\text{Max} \mid r-\rho \mid /f(x)(r^{\frac{1}{2}}+\rho^{\frac{1}{2}})$$

$$< \text{const. } [f(x)/x^{2/5}\delta_n][x^{3/5}f(x)/x],$$

by the mean-value theorem, $= O(x^{-1/5})$.

(C) Replacing α^2 by $(x-\xi)^2/4r\rho$ introduces an error of

$$< (kf(x)/x)\left(\int_{x_3}^{x_1}+\int_{x_2}^{x_4}\right)\int_0^{\frac{1}{2}\pi}[(x-\xi)^2/x^2][(x-\xi)^2/4r\rho+\sin^2\lambda]^{-3/2}d\lambda d\xi$$
$$= O[(f(x)/x)x^{3/5}x^{-2}f^3(x)/\delta_n] = O(x^{-\frac{1}{2}}).$$

(D) Replacing ρ by r in this same radical gives an error of

$$< [kf(x)/x]\left(\int_{x_3}^{x_1}+\int_{x_2}^{x_4}\right)\int_0^{\frac{1}{2}\pi}[(x-\xi)^2/4r] \mid 1/\rho - 1/r \mid$$
$$\times [(x-\xi)^2/4r\rho+\sin^2\lambda]^{-\frac{1}{2}}[(x-\xi)^2/4r^2+\sin^2\lambda]^{-\frac{1}{2}}$$
$$\times \{[(x-\xi)^2/4r\rho+\sin^2\lambda]^{\frac{1}{2}}+[(x-\xi)^2/4r^2+\sin^2\lambda]^{\frac{1}{2}}\}^{-1}d\lambda d\xi$$
$$< [kf(x)/x]\left(\int_{x_3}^{x_1}+\int_{x_2}^{x_4}\right)\int_0^{\frac{1}{2}\pi}[(x-\xi)^2/4r]x^{3/5}[1/x\,f(x)](\tfrac{1}{2})$$
$$\times [(x-\xi)^2/8r^2+\sin^2\lambda]^{-3/2}d\lambda d\xi$$

$= O(x^{-1/5})$, where k, as above, denotes a constant and where the mean-value theorem has again been used.

(E) Replacing $f(\xi)f'(\xi)$ by $f(x)f'(x)$ gives an error

$$< \left(\int_{x_3}^{x_1}+\int_{x_2}^{x_4}\right)[\mid f(\xi)f'(\xi)-f(x)f'(x)\mid /f(x)]$$
$$\int_0^{\frac{1}{2}\pi}2f(x)[(x-\xi)^2+4r^2\sin^2\lambda]^{-\frac{1}{2}}\,d\lambda d\xi.$$

Now for $\xi < x$, $\mid f(\xi)f'(\xi)-f(x)f'(x)\mid < f(x_3)f'(x_3)-f(x)f'(x)$

$$= [f^2(x_3)/2x_3 - f^2(x)/2x] - [f^2(x_3)/(4x_3 \log x_3) - f^2(x)/(4x \log x)]$$
$$+ \{[f^2(x_3) \log\log x_3]/(8x_3 \log^2 x_3) - [f^2(x) \log\log x]/(8x \log^2 x)\}$$
$$+ O(\log^{-5/2} x)$$

$= O(x^{-2/5}) + O(\log^{-5/2} x)$, where the mean value theorem is used to estimate the three differences, and (1) and (1.4) are used as usual. An almost identical calculation obtains the same result for $\xi > x$; accordingly the above error is

$$< \text{const.}\ (\log^{-5/2} x) \int_0^{\frac{1}{2}\pi} \left(\int_{x_3}^{x_1} + \int_{x_2}^{x_4} \right) [(x-\xi)^2 + 4r^2 \sin^2 \lambda]^{-\frac{1}{2}} d\xi d\lambda$$
$$= \text{const.}\ (\log^{-5/2} x) \int_0^{\frac{1}{2}\pi} \log\{[x^{3/5} + (x^{6/5} + 4r^2 \sin^2 \lambda)^{\frac{1}{2}}]/[\delta_n$$
$$+ (\delta^2{}_n + 4r^2 \sin^2 \lambda)^{\frac{1}{2}}]\} d\lambda$$
$$< \text{const.}\ (\log^{-5/2} x) \int_0^{\frac{1}{2}\pi} \log x \, d\lambda = O(\log^{-3/2} x).$$

At this point then we have:

$$R_2 = (-2/\pi) \left(\int_{x_3}^{x_1} + \int_{x_2}^{x_4} \right) f(x) f'(x) \int_0^{\frac{1}{2}\pi} [(x-\xi)^2 + 4r^2 \sin^2 \lambda]^{-\frac{1}{2}} d\lambda d\xi$$
$$+ O(\log^{-3/2} x)$$
$$= (-2/\pi) f(x) f'(x) \int_0^{\frac{1}{2}\pi} 2 \int_{\delta_n}^{x^{3/5}} (p^2 + 4r^2 \sin^2 \lambda)^{-\frac{1}{2}} dp d\lambda + O(\log^{-3/2} x)$$

after changing the order of integration and making a simple change of variable. Now integrating,

$$R_2 = (-4/\pi) f(x) f'(x) \int_0^{\frac{1}{2}\pi} \log\{[x^{3/5} + (x^{6/5} + 4r^2 \sin^2 \lambda)^{\frac{1}{2}}]/[\delta_n$$
$$+ (\delta^2{}_n + 4r^2 \sin^2 \lambda)^{\frac{1}{2}}]\} d\lambda + O(\log^{-3/2} x).$$

Since

$$\log[x^{3/5} + (x^{6/5} + 4r^2 \sin^2 \lambda)^{\frac{1}{2}}] = (3/5) \log x + \log 2 + O(x^{-1/5})$$

we find

$$R_2 = -2f(x) f'(x) [(3/5) \log x + \log 2]$$
$$+ (4/\pi) f(x) f'(x) \int_0^{\frac{1}{2}\pi} \log [\delta_n + (\delta^2{}_n + 4r^2 \sin^2 \lambda)^{\frac{1}{2}}] d\lambda + O(\log^{-3/2} x)$$
$$= -2f(x) f'(x) [(3/5) \log x + \log 2] + O(\log^{-3/2} x)$$
$$+ (4/\pi) f(x) f'(x) [\tfrac{1}{2}\pi \log 2f(x) + \int_0^{\frac{1}{2}\pi} \log(a + (a^2 + \sin^2 \lambda)^{\frac{1}{2}}) d\lambda],$$

where $a = \delta_n/2r = O(\log^{\frac{1}{2}-n} x)$ and so is small.

Consider $\int_0^{\frac{1}{2}\pi}[\log(a + (a^2 + \sin^2\lambda)^{\frac{1}{2}}) - \log\sin\lambda]\,d\lambda$. This tends to zero with a. However, a calculation involving the Maclaurin series in powers of a leads to a divergent integral; so we let $a = b^2$ and have by the mean value theorem:

$$| \log[b^2 + (b^4 + \sin^2\lambda)^{\frac{1}{2}}] - \log\sin\lambda | < | b | \operatorname{Max} | 2b/(b^4 + \sin^2\lambda)^{\frac{1}{2}} |$$
$$= | b | (2/\sin\lambda)^{\frac{1}{2}} = (2a/\sin\lambda)^{\frac{1}{2}},$$

so that

$$\int_0^{\frac{1}{2}\pi}[\log(a + (a^2 + \sin^2\lambda)^{\frac{1}{2}}) - \log\sin\lambda]\,d\lambda < \int_0^{\frac{1}{2}\pi}(2a/\sin\lambda)^{\frac{1}{2}}\,d\lambda$$
$$= O(a^{\frac{1}{2}}) = O(\log^{\frac{1}{4}-n/2} x),$$

or

$$\int_0^{\frac{1}{2}\pi}\log(a + (a^2 + \sin^2\lambda)^{\frac{1}{2}})\,d\lambda = \int_0^{\frac{1}{2}\pi}\log\sin\lambda\,d\lambda + O(\log^{\frac{1}{4}-n/2} x)$$
$$= (-\pi/2)\log 2 + O(\log^{\frac{1}{4}-n/2} x).$$

Using this in R_2 gives

$$R_2 = -2f(x)f'(x)[(3/5)\log x + \log 2]$$
$$+ (4/\pi)f(x)f'(x)[(\pi/2)\log(2r) - (\pi/2)\log 2] + O(\log^{-\frac{1}{4}-n/2} x)$$
$$+ O(\log^{-3/2} x),$$

or

$$R_2 = f(x)f'(x)[\log f^2(x) - (6/5)\log x - \log 4] + O(\log^{-\frac{1}{4}-n/2} x)$$
$$+ O(\log^{-3/2} x).$$

And now

$$R_3 = -\tfrac{1}{2\pi}\left(\int_{3x/4}^{x_3} + \int_{x_4}^{5x/4}\right) f(\xi)f'(\xi)[1 + f'^2(\xi)]^{-\frac{1}{2}}$$
$$\times \int_0^{\frac{1}{2}\pi} 4[(x-\xi)^2 + (r+\rho)^2 - 4r\rho\cos^2\lambda]^{-\frac{1}{2}}\,d\lambda\,d\xi.$$

In this last part of I_2 we may use the series for the elliptic integral so that, since $| x - \xi | \geq x^{3/5}$,

$$\int_0^{\frac{1}{2}\pi}[(x-\xi)^2 + (r+\rho)^2 - 4r\rho\cos^2\lambda]^{-\frac{1}{2}}4\,d\lambda$$
$$= 2\pi[(x-\xi)^2 + (r+\rho)^2]^{-\frac{1}{2}}[O(f^2(x)/x^{6/5}) + 1].$$

This gives

$$R_3 = -\left(\int_{3x/4}^{x_3} + \int_{x_4}^{5x/4}\right) f(\xi)f'(\xi)\,[1 + f'^2(\xi)]^{-\frac{1}{2}}[(x-\xi)^2 + (r+\rho)^2]^{-\frac{1}{2}} \times [1 + O(x^{-1/5})]\,d\xi.$$

The error term contributes

$$< \text{const.}\ [f^2(x)/x][f^2(x)/x^{6/5}] \int_{x_4}^{5x/4} d\xi/|\,x-\xi\,| = O(x^{-1/5}).$$

To simplify the other integrals note:

(A) Replacing $[1 + f'^2(\xi)]^{-\frac{1}{2}}$ by 1 gives an error of

$$< \text{const.}\ x[f^2(x)/x^2]x^{-3/5} = O(x^{-3/5}).$$

(B) Replacing $[(x-\xi)^2 + (r+\xi)^2]^{-\frac{1}{2}}$ by $|\,x-\xi\,|^{-1}$ gives an error of

$$< \text{const.}\ [f^2(x)/x][f^2(x)/x^{6/5}] \int_{x_4}^{5x/4} d\xi/|\,x-\xi\,| = O(x^{-1/5}).$$

With these changes

$$R_3 = -\left(\int_{3x/4}^{x_3} + \int_{x_4}^{5x/4}\right) f(\xi)f'(\xi)\,|\,x-\xi\,|^{-1}d\xi + O(x^{-1/5}).$$

Using (1), this becomes

$$R_3 = -\left(\int_{3x/4}^{x_3} + \int_{x_4}^{5x/4}\right)[f^2(\xi)/2\xi][1 - (1/(2\log\xi)) + (\log\log\xi)/(4\log^2\xi)]\,d\xi/|\,\xi - x\,| + O(\log^{-3/2}x).$$

Now for $\xi > x$,

$$|\,f^2(x)/x - f^2(\xi)/\xi\,| \leq f^2(x)/x - f^2(5x/4)/(5x/4) = O(\log^{-3/2}x),$$

where we use (1. 4) exactly as before; for $\xi < x$, the argument is again almost the same and the result identical. If then in R_3 we replace $f^2(\xi)/\xi$ by $f^2(x)/x$, the error introduced into the first term is not small enough, but in the last two terms it is

$$< \text{const.}\ \log^{-5/2}x \int_{x_4}^{5x/4} d\xi/(\xi - x) = O(\log^{-3/2}x).$$

Again using identities very similar to A, B, and C we find almost at once that replacing $1/\log\xi$ by $1/\log x$ and $\log\log\xi/\log^2\xi$ by $\log\log x/\log^2 x$ introduces errors which are $O(\log^{-3/2}x)$. Accordingly,

$$R_3 = -\left(\int_{3x/4}^{x_3} + \int_{x_4}^{5x/4}\right)\{[f^2(\xi)/2\xi] + [f^2(x)/2x][(-1/(2\log x)) + (\log\log x)/(4\log^2 x)]\}(d\xi/|\,\xi - x\,| + O(\log^{-3/2}x)$$

$$= -\left(\int_{3x/4}^{x_3} + \int_{x_4}^{5x/4}\right) (f^2(\xi)/2\xi)\, d\xi/|\,\xi - x\,|$$
$$- [f^2(x)/2x][-2/5 + (\log\log x)/(5\log x)] + O(\log^{-3/2} x).$$

Calling the remaining integrals R_4,

$$R_4 = -\left(\int_{3x/4}^{x_3} + \int_{x_4}^{5x/4}\right)[f^2(x)/2x]\, d\xi/|\,\xi - x\,|$$
$$+\left(\int_{3x/4}^{x_3} + \int_{x_4}^{5x/4}\right)[f^2(x)/2x - f^2(\xi)/2\xi]\, d\xi/|\,\xi - x\,|$$
$$= -[f^2(x)/2x][(4/5)\log x - 2\log 4]$$
$$-\left(\int_{3x/4}^{x_3} + \int_{x_4}^{5x/4}\right)\int_x^{\xi}[(f(t)f'(t)/t) - (f^2(t)/2t^2)]\, dt\, d\xi/|\,\xi - x\,|$$
$$= -[f^2(x)/2x][(4/5)\log x - 2\log 4]$$
$$-\left(\int_x^{x_4}\int_{x_4}^{5x/4} + \int_{x_4}^{5x/4}\int_t^{5x/4} + \int_x^{x_3}\int_{3x/4}^{x_3} + \int_{x_3}^{3x/4}\int_{3x/4}^{t}\right)$$
$$[(f(t)f'(t)/t) - (f^2(t)/2t^2)]\, d\xi\, dt/|\,\xi - x\,|.$$

Now by (1),

$$(f(t)f'(t)/t) - (f^2(t)/2t^2)$$
$$= [f^2(t)/2t^2][(-1/(2\log t)) + (\log\log t)/(4\log^2 t) + O(\log^{-2} t)].$$

Since the first and third integrals have constant limits they can be calculated directly and are $O(\log^{-3/2} x)$, so that

$$R_4 = -[f^2(x)/2x][(4/5)\log x - 2\log 4]$$
$$-\left(\int_{x_4}^{5x/4}\int_t^{5x/4} + \int_{x_3}^{3x/4}\int_{3x/4}^{t}\right)[f^2(t)/2t^2][(-1/(2\log t))$$
$$+ (\log\log t)/(4\log^2 t)]\, d\xi\, dt/|\,\xi - x\,| + O(\log^{-3/2} x)$$
$$= -[f^2(x)/2x][(4/5)\log x - 2\log 4]$$
$$-\left(\int_{x_4}^{5x/4} + \int_{x_3}^{3x/4}\right)[f^2(t)/2t^2][(-1/(2\log t))$$
$$+ (\log\log t)/(4\log^2 t)]\log[(x/4)/|\,x - t\,|]\, dt + O(\log^{-3/2} x).$$

Again as before, for $t > x$,

$$
\begin{aligned}
&|(f^2(x)/x\log x) - (f^2(t)/t\log t)| \\
&\qquad < [f^2(x)/x\log x] - [f^2(5x/4)/(5x/4)\log(5x/4)] \\
&\qquad = O(\log^{-5/2} x);
\end{aligned}
$$

similarly for $t < x$. Also
$|[f^2(x)\log\log x/x\log^2 x] - [f^2(t)(\log\log t)/(t\log^2 t)]| = O(\log^{-5/2} x)$
both for $t > x$ and $t < x$. In R_4 these differences contribute $O(\log^{-3/2} x)$ so that

$$
\begin{aligned}
R_4 = &- [f^2(x)/2x][(4/5)\log x - 2\log 4] \\
&+ \{[f^2(x)/(4x\log x)] - [f^2(x)(\log\log x)/(8x\log^2 x)]\} \\
&\left(\int_{x_4}^{5x/4} + \int_{x_3}^{3x/4}\right)\log[(x/4)/|x - t|]t^{-1}dt + O(\log^{-3/2} x) \\
= &- [f^2(x)/2x][(4/5)\log x - 2\log 4] + O(\log^{-3/2} x),
\end{aligned}
$$

since direct evaluation of the remaining integrals shows them to be $O(1)$.

In summary then

$$
\begin{aligned}
I_2 = &\int_{3x/4}^{x} f(\xi)f'(\xi)/(\xi + 1)\,d\xi \\
&+ [f^2(x)/2x][\log(5/4) + 2/5 - (\log\log x)/(5\log x) \\
&\qquad - (4/5)\log x + 2\log 4] \\
&+ f(x)f'(x)[\log f^2(x) - (6/5)\log x - \log 4] + O(\log^{-3/2} x) \\
= &\int_{3x/4}^{x} f(\xi)f'(\xi)/(\xi + 1)\,d\xi \\
&+ [f^2(x)/2x][\log(f^2(x)/x^2) + \log 5 + \tfrac{1}{2}] + O(\log^{-3/2} x).
\end{aligned}
$$

Next we consider

$$
I_3 = -\tfrac{1}{2\pi}\int\int_{W_3}(1/R - 1/1 + \xi)\phi_n dS
$$

$$
\begin{aligned}
1/R &= [(x - \xi)^2 + r^2 + \rho^2 - 2r\rho\cos(\theta - \lambda)]^{-\frac{1}{2}} \\
&= (1/\xi - x)[1 + O(f^2(\xi)/\xi^2)]. \\
&= 1/(\xi - x) + O(1/\xi^2).
\end{aligned}
$$

So

$$I_3 = -\tfrac{1}{2\pi}\int\int_{W_3}[(x+1)/(\xi-x)(\xi+1)+O(1/\xi^2)][f'(\xi)+O(f'^3(\xi))]dS$$

$$= -\int_{5x/4}^{\infty}[(x+1)/(\xi-x)(\xi+1)]f(\xi)f'(\xi)d\xi + O(\int_{5x/4}^{\infty}(f^2(\xi)/\xi)d\xi/\xi^2)$$

$$= -\int_{5x/4}^{\infty}[(x+1)/(\xi-x)(\xi+1)]f(\xi)f'(\xi)d\xi + O(1/x).$$

Replacing $x+1$ by x and $\xi+1$ by ξ both give errors less than

$\int_{5x/4}^{\infty} d\xi/\xi^2 = O(1/x)$, so that using (1),

$$I_3 = -\int_{5x/4}^{\infty}[x/\xi(\xi-x)]f(\xi)f'(\xi)\,d\xi + O(1/x)$$

$$= -\int_{5x/4}^{\infty}[x/\xi(\xi-x)][f^2(\xi)/2\xi]d\xi + O(\log^{-3/2}x),$$

since

$$x\int_{5x/4}^{\infty}1/\xi(\xi-x)\,d\xi < x\int_{5x/4}^{\infty}5/\xi^2\,d\xi = 4.$$

Now

$$\int_{5x/4}^{\infty}f^2(\xi)\,d\xi/2\xi^2(\xi-x) = \int_{5x/4}^{\infty}f^2(x)\,d\xi/2x\xi(\xi-x)$$

$$+\int_{5x/4}^{\infty}\int_{x}^{\xi}[f(t)f'(t)/t-f^2(t)/2t^2]dt\,d\xi/\xi(\xi-x)$$

$$= [f^2(x)/2x][(\log 5)/x]$$

$$\left(\int_{x}^{5x/4}\int_{5x/4}^{\infty}+\int_{5x/4}^{\infty}\int_{t}^{\infty}\right)O(1)/(2t\log^{3/2}t)\,d\xi dt/\xi(\xi-x),$$

where we have used (1) and (1.4) and changed the order of integration. Of the two remaining integrals the first is

$$< \tfrac{1}{2}\log^{-3/2}x\int_{x}^{\infty}\int_{5x/4}^{\infty}O(1)\,(5d\xi/\xi^2)\,dt/t = O(1/x\log^{-3/2}x),$$

and the second is

$$< \tfrac{1}{2}\log^{-3/2}x\int_{5x/4}^{\infty}\int_{t}^{\infty}O(1)\,(5d\xi/\xi^2)\,dt/t = O(1/x\log^{-3/2}x);$$

and so $I_3 = -\log 5\,f^2(x)/2x + O(\log^{-3/2}x)$.

Finally we treat $J_2 = \frac{1}{2\pi}\int\int_{W_2} \phi(\xi, \rho) \cos\psi/R^2 dS$. Using (6. 5),

$$J_2 = \tfrac{1}{2\pi}[C_2 + \tfrac{1}{2}B(x)] \int\int_{W_2} \cos\psi/R^2\, dS$$
$$-(1/4\pi)\int\int_{W_2}[B(x) - B(\xi)] \cos\psi/R^2\, dS + O(x^{-\frac{1}{2}}),$$

and now by (3. 15)

$$J_2 = -C_2 - \tfrac{1}{2}B(x) + O(B(x)/x) + S_1 + O(x^{-\frac{1}{2}}),$$

where S_1 is the remaining integral. By (1) and (1. 4)

$$B(x) = \int_a^x f'^2(t)\,dt < \text{const.} \int_a^x 1/(t \log^{\frac{1}{2}} t)\,dt = O(\log x),$$

so that the first error term in J_2 above is $O(x^{-\frac{1}{2}})$.

Now let W_{21} be that portion of W_2 for which ξ is between $x - b$ and $x + b$, and let W_{22} be the remaining two portions. Then in W_{21}

$$|\,B(x) - B(\xi)\,| = |\int_\xi^x f'^2(t)\,dt\,| \leq |\int_\xi^x c^2/(t\log^{\frac{1}{2}} t)\,dt\,|$$

$= O[b/(x \log^{\frac{1}{2}} x)] = O(\log^{-n} x)$, so that from this region S_1 receives a contribution of

$$< k \log^{-n} x \int\int_{W_{21}} - \cos\psi/R^2\, dS < k \log^{-n} x \int\int_{W_2} - \cos\psi/R^2\, dS$$
$$= O(\log^{-n} x),$$

k being a constant and n arbitrary.

Furthermore, since $B(x) = O(\log x)$, the contribution to S_1 from the region W_{22} is

$$< k \log x \int\int_{W_{22}} f(\xi)\, b^{-2} d\xi d\lambda = O(x \log^{2n+1} x/x^{3/2}) = O(x^{-1/3}).$$

Consequently $J_2 = -C_2 - \frac{1}{2}B(x) + O(\log^{-n} x)$.

Part II. Calculation of New Terms.

Inserting (6. 5), (6. 6), (6. 7) and our reappraisals of I_1, I_2, I_3, and J_2 into (3. 10) we have

$$B(x) = \int_a^x f(\xi)f'(\xi)/(\xi + 1)\,d\xi + \text{const.} + [f^2(x)/x] \log[f(x)/x]$$
$$+ \tfrac{1}{2}dc^2[\log^{-\frac{1}{2}} x - (\log\log x)/(4\log^{3/2} x)] + O(\log^{-3/2} x),$$

where $d = \frac{1}{2} - \log 4$; now by using (6. 15) and (6. 17) this becomes

$$E_1 - F_1 = (1/4) \int_a^x f^2(\xi)/\xi^2\, d\xi - [f^2(x)/x] \log [x/f(x)] - K/4$$
$$+ \tfrac{1}{2} dc^2 [\log^{-\frac{1}{2}} x - (\log \log x)/(4 \log^{3/2} x)] + O(\log^{-3/2} x),$$

where K is a constant. By (6. 15) and (6. 17)

$$E_1 - F_1 = \int_x^\infty [f(\xi) f'(\xi)/(\xi + 1) - f^2(\xi)/4\xi^2 - f'^2(\xi)] d\xi.$$

As usual, replacing $\xi + 1$ by ξ gives an error of $O(1/x)$, so using (1)

$$E_1 - F_1 = \int_x^\infty O(f^2(\xi)/\xi^2 \log^2 \xi)\, d\xi + O(1/x) = O(\log^{-3/2} x).$$

Accordingly, letting $f(x) = x^{\frac{1}{2}} h(x)/\log^{1/4} x$, we have $\int_a^x h^2(\xi)/\xi \log^{\frac{1}{2}} \xi\, d\xi$

$$= 4h^2(x) \log^{-\frac{1}{2}} x [\tfrac{1}{2} \log x + (1/4) \log \log x - \log h(x)] + K$$
$$- 2dc^2 [\log^{-\frac{1}{2}} x - (\log \log x)/(4 \log^{3/2} x)] + O(\log^{-3/2} x).$$

Substituting $y = \log x$, $h^2(x) = H(y)$, $b = \log a$, and using (6. 19), this becomes

$$R(y) = K + y^{-\frac{1}{2}} (\log y) H(y) - 2y^{-\frac{1}{2}} H(y) \log H(y)$$
$$- 2dc^2 [y^{-\frac{1}{2}} - y^{-3/2} (\log y)/4] + O(y^{-3/2}),$$

and so by (1. 4)

$$R(y) = K + O(y^{-3/2+\epsilon}) + c^2 y^{-\frac{1}{2}} \log y - (4c^2 \log c + 2dc^2) y^{-\frac{1}{2}}.$$

Using this expression for $R(y)$ in (6. 20), carrying out the integrations and combining terms, we find:

$$H(y) = c^2 [1 - \log y/4y + (\log c + 1/4 + d/2)/y] + O(y^{-2+\epsilon}).$$

Returning to our first expression for $R(y)$ above, and using this latest value for $H(y)$, direct substitution followed by recombination of terms gives:

$$R(y) = K + c^2 y^{-\frac{1}{2}} \log y - 2c^2 (2 \log c + d) y^{-\frac{1}{2}} - c^2 y^{-3/2} (\log^2 y)/4$$
$$+ c^2 (2 \log c + 3/4 + d) y^{-3/2} \log y + O(y^{-3/2}).$$

Now substituting this value of $R(y)$ into (6. 20), carrying out the integrations and combining terms, we find

$$H(y) = c^2 [1 - \log y/4y + (\log c + 1/4 + d/2)/y]$$
$$+ 3 \log^2 y/32y^2 - [3 (\log c)/4 + 5/16 + 3d/8] \log y/y^2 + O(1/y^2).$$

Using the binomial theorem and recombining,

$$(H(y))^{\frac{1}{2}} = h(x) = c[1 - \log y/8y + ((\log c)/2 + 1/8 + d/4)/y$$

$$+ 5\log^2 y/128y^2 - (5(\log c)/16 + 9/64 + 5d/32)\log y/y^2 + O(1/y^2)].$$

Finally, recalling that $y = \log x$ and that $f(x) = x^{\frac{1}{2}}h(x)/\log^{1/4} x$ we have for $f(x)$ the expression written at length in the introduction.

The author wishes to express his thanks to Professor Levinson for his interest and suggestions during the preparation of this paper.

BOSTON UNIVERSITY.

REFERENCES.

[1] Norman Levinson, "On the asymptotic shape of the cavity behind an axially symmetric nose moving through an ideal fluid," *Annals of Mathematics*, vol. 47 (1946), pp. 704-730.

[2] Decimal equation numbers are as in the paper of Levinson.

ON ORDINARY DIFFERENTIAL EQUATIONS OF ANY EVEN ORDER AND THE CORRESPONDING EIGENFUNCTION EXPANSIONS.*

By Kunihiko Kodaira.

The general theory of expanding an arbitrary function in terms of the eigenfunctions of a singular differential operator of the second order was first given by H. Weyl.[1] An alternative method, based on the general theory of linear transformations in the Hilbert space, is to be found in a treatise by M. H. Stone.[2] Recently E. C. Titchmarsh has treated the same theory by still another method and obtained some new results of importance for applications.[3] The purpose of the present paper is to generalize the Weyl-Stone-Titchmarsh theory of eigenfunction expansions to the case of formally self-adjoint ordinary differential operators with real coefficients of any even order. Our method is also based on the general theory of linear transformations in the Hilbert space and goes along the lines of H. Weyl.[4] In Section **1** we shall give preliminary remarks on the formally self-adjoint differential operator of real coefficients, which will be denoted by L. In Section **2** we shall analyse the singularities of the operator L. For that purpose, we introduce a complex projective space $\mathfrak{P}$, each point of which is associated with a general solution of the differential equation $L[u] = l \cdot u$, and construct the subsets $\mathfrak{f}_a(l)$, $\mathfrak{f}_b(l)$ of $\mathfrak{P}$ corresponding to the "limit circles" in the theory of H. Weyl.[5] Furthermore it will be shown that the formally self-adjoint differential operators of the even order n can be classified into $(\frac{1}{2}n + 1)^2$ different types according to the nature of their singularities. In Section **3** we shall introduce Green's function and determine the closure and the adjoint of the operator L. The boundary conditions will be discussed in Section **4**. It will be shown that, under suitable boundary conditions, L becomes a self-adjoint operator. The next section, **5**, will be devoted to the proof of the spectral theorem, which can be regarded as a generalization of the Weyl-Stone-Titchmarsh results. In

---

* Received September 22, 1949.

[1] Weyl, [10], [11], [12]. Numbers in brackets refer to the bibliography at the end of the paper.

[2] Stone, [7], Chap. X, §3.

[3] Titchmarsh, [9]. Another proof of Titchmarsh's results is given in Kodaira [3].

[4] Weyl, [11], [12].

[5] Weyl, [11], pp. 221-231.

Section **6**, the expansion theorem will be deduced from the spectral theorem. Finally it will be shown in Section **7** that the spectral and the expansion theorems can be readily extended to the case of simultaneous differential equations.

The author wishes to express his grateful thanks to Prof. H. Weyl for his encouragement and his aid in acquainting the author with the recent results of Prof. E. C. Titchmarsh.

1. Introduction. Consider a formal differential operator

$$L = p_0(x)(d/dx)^n + p_1(x)(d/dx)^{n-1} + \cdots + p_n(x), \qquad (a < x < b)$$

of the even order $n = 2\nu$ defined in a (finite or infinite) *open* interval (a, b), where each coefficient $p_m(x)$ $(m = 0, 1, \cdots, n)$ is a *real valued continuous function defined in* (a, b) *having continuous derivatives up to the order* $n - m$ *and* $p_0(x) > 0$ [or $p_0(x) < 0$]; for $x \rightarrow a$ or $x \rightarrow b$, $p_m(x)$ may behave arbitrarily, e. g. increase infinitely, oscillate infinitely many times.[6] Furthermore we assume that L is *formally self-adjoint,* i. e. L coincides with its Lagrange adjoint or

$$(1.1)_m \quad p_m(x) = \sum_{k=0}^{m-1} (-1)^k C_{m-k}^{n-k} p_k^{(m-k)}(x) + (-1)^m p_m(x), \qquad (m = 1, 2, \cdots, n),$$

where C_r^s mean binomial coefficients. For *odd* m, $(1.1)_m$ is rewritten as

$$(1.2)_m \qquad 2p_m(x) = C_m^n p_0^{(m)}(x) - C_{m-1}^{n-1} p_1^{(m-1)}(x) + - \cdots + C_1^{n-m+1} p'_{m-1}(x), \qquad (m = 1, 3, 5, \cdots, n-1),$$

while, *for even* m, $(1.1)_m$ *follows from* $(1.1)_1, (1.1)_2, \cdots, (1.1)_{m-1}$. Hence, in general, L *is formally self adjoint if the coefficients* $p_0(x), \cdots, p_{n-1}(x)$ *satisfy the conditions* $(1.2)_m$, $m = 1, 3, \cdots, n-1$. Thus, in our operator L, *the coefficients* $p_0(x), p_2(x), p_4(x), \cdots, p_n(x)$ *may be chosen arbitrarily, while* $p_1(x), p_3(x), \cdots$ *must be determined successively by* $(1.2)_1, (1.2)_3, \cdots, (1.2)_{n-1}$.

Put now

$$L[u] = p_0(x)(d^n u/dx^n) + p_1(x)(d^{n-1}u/dx^{n-1}) + \cdots + p_n(x) \cdot u$$

for arbitrary functions $u(x)$ having continuous derivatives up to the order n. Then we have *Green's formula*

---

[6] These conditions on the coefficients may be replaced by weaker ones, but in the present paper we are not interested with such generalizations. Cf. Stone, [7], pp. 448-453, Halperin, [2].

$$(1.3) \qquad \int_y^x (L[u]\cdot v - u\cdot L[v])\,dx = [uv](x) - [uv](y),$$

where $[uv](x)$ means the *bilinear form*

$$[uv](x) = \sum_{j+k\leqq n-1} B_{jk}(x)u^{(j)}(x)v^{(k)}(x)$$

of two *"vectors"* $(u, u', u'', \cdots, u^{(n-1)})$, $(v, v', v'', \cdots, v^{(n-1)})$ with the coefficients

$$(1.4) \qquad B_{jk}(x) = (-1)^{j+1} \sum_{h=0}^{n-j-k-1} (-1)^h C_k^{n-j-1-h} p_h^{(n-j-1-k-h)}.$$

The bilinear form $[uv](x)$ *is obviously skew-symmetric,* i. e.

$$(1.5) \qquad [uv](x) = -[vu](x).^{7}$$

For $j+k=n-1$, we have

$$(1.6) \qquad B_{jk}(x) = (-1)^{j+1}p_0(x), \qquad (j+k=n-1).$$

The bilinear form $[uv](x)$ *is therefore non-degenerate,* since by hypothesis $p_0(x)\neq 0$. Now we introduce the *"Wronskian"*

$$[u_1u_2\cdots u_n](x) = \{-p_0(x)\}^{\nu}\cdot \det(u_j^{(k-1)}), \qquad (j, k = 1, 2, \cdots, n).$$

Then we have the identity

$$(1.7) \qquad (1/2^{\nu}\nu!)\Sigma \pm [u_1u_{\nu+1}][u_2u_{\nu+2}]\cdots[u_{\nu}u_n] = [u_1u_2\cdots u_n],$$

where $\Sigma\pm$ means the alternating sum extending to all $n!$ permutations of n functions $u_1, u_2, \cdots, u_n$. (1.7) is proved as follows:[8] As one readily infers, the left hand side of (1.7) is equal to

$$[u_1u_2\cdots u_n]\cdot(-1)^{\nu(\nu+1)/2}p_0^{-\nu}(1/2^{\nu}\nu!)\Sigma\pm B_{01}B_{23}\cdots B_{n-2\; n-1},$$

$\Sigma\pm$ being the alternating sum extending to all $n!$ permutations of the suffixes $0, 1, 2, \cdots, n-1$; while, since $B_{jk}=0$ for $j+k\geqq n$, we get, using (1.6),

$$(1/2^{\nu}\nu!)\Sigma\pm B_{01}B_{23}\cdots B_{n-2\; n-1} = B_{0\; n-1}B_{1\; n-2}\cdots B_{\nu-1\; \nu} = (-1)^{\nu(\nu+1)/2}p_0^{\nu}.$$

Hence we have (1.7).

2. Classification. We shall consider the *differential equation*

$$(2.1) \qquad L[u] = l\cdot u,$$

where l means a complex parameter. From (1.3) follows that, *if u and v*

---

[7] This identity can be deduced readily from Green's formula.

[8] Weyl, [14], pp. 166-167.

satisfy one and the same equation (2.1), $[uv](x)$ *does not depend on* x. Combined with (1.7), this shows further that *the Wronskian* $[u_1u_2\cdots u_n](x)$ *of* n *solutions* $u_1(x), u_2(x), \cdots, u_n(x)$ of the *equation* (2.1) *is a constant.* In these cases, we write therefore $[uv]$ or $[u_1u_2\cdots u_n]$ for $[uv](x)$ or $[u_1\cdots u_n](x)$, respectively.

DEFINITION 2.1. *By a regular system of fundamental solutions of the differential equation* $L[u] = l\cdot u$ *we shall mean a system of* n *solutions* $s_j(x, l)$ $(j = 1, 2, \cdots, n)$ *such that*

$$(2.2) \qquad [s_1s_2\cdots s_n] = 1, \qquad\qquad (2.3) \qquad s_j(x, \bar{l}) = \bar{s}_j(x, l),$$

and that the functions $d^ms_j(x, l)/dx^m$ $(j = 1, 2, \cdots, n;\ m = 0, 1, \cdots, n-1)$ *are entire functions of* l. *A regular system of fundamental solutions will be called a canonical system, if the skew symmetric matrix* $([s_js_k])$ *has the " canonical form " in the sense that*

$$(2.4) \qquad [s_js_k] = \epsilon_{jk},$$

ϵ_{jk} *being the quantities defined by* $\epsilon_{jk} = 1$ *(for* $k = j + \nu$*)*, $= -1$ *(for* $k = j - \nu$*)*, $= 0$ *(otherwise).*

It should be noted here that (2.4) *implies* (2.2), as (1.7) shows.

A canonical system of fundamental solutions is obtained, for example, as follows: Choose a fixed point c, $a < c < b$, arbitrarily (in what follows c means always a fixed point such that $a < c < b$). Then there exist n *real* vectors

$$e_j = (e_j, e'_j, e''_j, \cdots, e_j^{(n-1)}), \qquad (j = 1, 2, \cdots, n)$$

such that $[e_je_k](c) = \epsilon_{jk}$. Now, under the boundary conditions

$$[d^mu/dx^m]_{x=c} = e_j^{(m)}, \qquad (m = 0, 1, \cdots, n-1),$$

the differential equation $L[u] = l\cdot u$ has one and only one solution, which will be denoted by $s_j^0(x, l)$. As one readily infers, the functions $s_j^0(x, l)$ $(j = 1, 2, \cdots, n)$ thus defined constitute a canonical system.

Obviously an arbitrary regular system of fundamental solutions $s_1, s_2, \cdots, s_n$ is obtained from the canonical system defined above by a unimodular transformation

$$(2.5) \qquad s_j(x, l) = \sum_{k=1}^{n} U_j^k(l)s_k^0(x, l), \qquad (\det(U_j^k) = 1),$$

where $U_j^k(l)$ $(j, k = 1, \cdots, n)$ are entire functions of l such that $U_j^k(\bar{l}) = \bar{U}_j^k(l)$.

Suppose now a regular system of fundamental solutions $s_j(x, l)$ $(j = 1,$

$2, \cdots, n$) as given. Then every solution $w(x)$ of the differential equation $L[w] = l \cdot w$ is represented as

$$w(x) = f^1 s_1(x, l) + f^2 s_2(x, l) + \cdots + f^n s_n(x, l).$$

We associate with such $w(x)$ the vector $f = (f^1, \cdots, f^n)$ and denote $w(x)$ by $w(x, l, f)$. For the given solution $w(x) = w(x, l, f)$, the vector $f = (f^1, \cdots, f^n)$ depends clearly on the choice of the fundamental solutions $s_1, s_2, \cdots, s_n$: *under the unimodular transformation*

$$\text{(U)} \qquad \tilde{s}_j(x, l) = \Sigma\, U_j^k(l) s_k(x, l), \qquad (\det(U_j^k) = 1)$$

$U_j^k(l)$ being entire functions of l such that $U_j^k(\bar{l}) = \bar{U}_j^k(l)$, f^j are transformed contragrediently to s_j, i. e.

$$\text{(V)} \qquad \tilde{f}^j = \Sigma\, V_k^j(l) f^k,$$

where $(V_k^j(l))$ means the inverse matrix of $(U_j^k(l))$. Obviously *$V_k^j(l)$ are also entire functions of l and satisfy $V_k^j(\bar{l}) = \bar{V}_k^j(l)$.*

Now we introduce the *$(n-1)$-dimensional projective space $\mathfrak{P}$ and consider $(f^1, f^2, \cdots, f^n)$ (excluding $(0, 0, \cdots, 0)$) as the homogeneous coordinates of a point in $\mathfrak{P}$ which will be denoted by (f).* If there is no possibility of confusion, *we write simply f for (f)*; especially, for an arbitrary subset $\mathfrak{S}$ of $\mathfrak{P}$, *we write $f \in \mathfrak{S}$ for $(f) \in \mathfrak{S}$.*

Choose a fixed point c, $a < c < b$, arbitrarily and, for arbitrary l with $\Im l \neq 0$, *denote by $\mathfrak{m}_a(l)$ or $\mathfrak{m}_b(l)$ the linear subspace of $\mathfrak{P}$ consisting of all points (f) such that $w(x, l, f)$ is square summable in $(a, c]$ or $[c, b)$, respectively.* Since it does not depend on the choice of c, $a < c < b$, whether $w(x, l, f)$ is square summable in $(a, c]$ (or $[c, b)$) or not, *$\mathfrak{m}_a(l)$ and $\mathfrak{m}_b(l)$ are independent of the choice of c. The linear spaces $\mathfrak{m}_a(l)$, $\mathfrak{m}_b(l)$ will be called main spaces with respect to the operator L.* From (2.3) follows

$$(2.6) \qquad w(x, \bar{l}, \bar{f}) = \bar{w}(x, l, f),$$

whence we get

$$(2.7) \qquad \mathfrak{m}_a(\bar{l}) = \bar{\mathfrak{m}}_a(l), \qquad \mathfrak{m}_b(\bar{l}) = \bar{\mathfrak{m}}_b(l).$$

For arbitrary l with $\Im l \neq 0$, we put

$$h(f; x, l) = [w\bar{w}](x)/2i\Im l, \qquad (w = w(x, l, f)).$$

Obviously $h(f; x, l)$ is a hermitian form in f and, because of the Green's formula, we have

$$(2.8)\qquad h(f;x,l)-h(f;y,l)=\int_y^x|w(x,l,f)|^2\,dx,\qquad (\Im l\neq 0).$$

Hence, for fixed f and l, *$h(f;x,l)$ is a monotone increasing function of x.* Now we divide the space $\mathfrak{P}$ into three parts:

$$\mathfrak{P}^+(x,l)=\{(f);h(f;x,l)>0\},$$
$$\mathfrak{P}^0(x,l)=\{(f);h(f;x,l)=0\},$$
$$\mathfrak{P}^-(x,l)=\{(f);h(f;x,l)<0\}.$$

Evidently $\mathfrak{P}^+(x,l)$ and $\mathfrak{P}^-(x,l)$ are open subsets of $\mathfrak{P}$; $\mathfrak{P}^0(x,l)$ is a quadratic hypersurface, constituting the common boundary of $\mathfrak{P}^+(x,l)$ and $\mathfrak{P}^-(x,l)$. Denote in general the closure of a subset $\mathfrak{s}$ in $\mathfrak{P}$ by $[\mathfrak{s}]$. Since $h(f;x,l)$ is a monotone increasing function of x, $\mathfrak{P}^-(x,l)$ is monotone decreasing in the sense that $[\mathfrak{P}^-(x,l)]\subset\mathfrak{P}^-(y,l)$ for $x>y$; similarly we have $[\mathfrak{P}^+(y,l)]\subset\mathfrak{P}^+(x,l)$ for $x>y$. Putting $\mathfrak{k}_b(l)=\bigcap_x\mathfrak{P}^-(x,l)$, $\mathfrak{k}_a(l)=\bigcap_x\mathfrak{P}^+(x,l)$, we infer therefore that *$\mathfrak{k}_a(l)$ and $\mathfrak{k}_b(l)$ are compact, not empty, and have no common points.*[9] *Furthermore we have*

$$(2.9)\qquad \mathfrak{k}_a(l)\subset\mathfrak{m}_a(l),\qquad \mathfrak{k}_b(l)\subset\mathfrak{m}_b(l).$$

Indeed, if $f\,\varepsilon\,\mathfrak{k}_b(l)$, then f lies in all $\mathfrak{P}^-(x,l)$, $a<x<b$, and therefore, by (2.8),

$$\int_c^x|w(x,l,f)|^2\,dx+h(f;c,l)=h(f;x,l)<0.$$

This shows clearly that $w(x,l,f)$ is square summable in $[c,b)$, proving that f is contained in $\mathfrak{m}_b(l)$; thus we get the second formula in (2.9). The first formula can be proved similarly. Again, from (2.6) follows

$$(2.10)\qquad h(f;x,\bar{l})=h(f;x,l).$$

This yields immediately $\mathfrak{P}^\pm(x,\bar{l})=\bar{\mathfrak{P}}^\pm(x,l)$ and consequently

$$(2.11)\qquad \mathfrak{k}_a(\bar{l})=\bar{\mathfrak{k}}_a(l),\qquad \mathfrak{k}_b(\bar{l})=\bar{\mathfrak{k}}_b(l).$$

Under the unimodular transformation (U) of the fundamental solutions mentioned above, the function $h(f;x,l)$ is invariant, whereas the sets $\mathfrak{P}^\pm(x,l)$ and their common parts $\mathfrak{k}_a(l)$, $\mathfrak{k}_b(l)$ are transformed according to the transformation (V) which is contragredient to (U).

For our purpose it is convenient to introduce the "*skew-product*"

$$[fg]_l=[w(l,f)w(l,g)]=\Sigma[s_j(l)s_k(l)]\cdot f^jg^k.$$

---

[9] The sets $\mathfrak{k}_a(l)$, $\mathfrak{k}_b(l)$ correspond to the "limit circle" in the theory of H. Weyl. See Weyl, [11], pp. 221-231; [13], pp. 235-238.

DEFINITION 2.2. *By a null plane with respect to l will be meant a linear subspace $\mathfrak{p}$ of $\mathfrak{P}$ such that $[fg]_l = 0$ for all $f, g \in \mathfrak{p}$.*

In case the fundamental solutions $s_1, s_2, \cdots, s_n$ constitute a canonical system, the skew product becomes

$$[fg]_l = \sum_{j=1}^{\nu} (f^j g^{\nu+j} - f^{\nu+j} g^j),$$

so that, in this case, *the notion of "null plane" does not depend on l.*

THEOREM 2.1. *The sets $\mathfrak{k}_a(l)$, $\mathfrak{k}_b(l)$ contain respectively at least one $(\nu - 1)$-dimensional null plane with respect to l.*

To show this we shall first prove

LEMMA 2.1. *Every $[\mathfrak{P}^+(y, l)]$, $a < y < b$, contains at least one $(\nu - 1)$-dimensional null plane $\mathfrak{p}$.*

Proof. As the system of fundamental solutions we may choose without loss of generality the canonical system constituted of the solutions $s_j^0(x, l)$ $(j = 1, 2, \cdots, n)$ satisfying the *real* boundary conditions $[d^m s_j^0(x, l)/dx^m]_{x=y} = e_j^{(m)}$, $(m = 0, 1, \cdots, n - 1; j = 1, \cdots, n)$ at the point y. Then, introducing a new system of coordinates:

$$\bar{f}^j = (f^j + i f^{\nu+j})/\sqrt{2}, \qquad \bar{f}^{\nu+j} = (f^j - i f^{\nu+j})/\sqrt{2}, \qquad (j = 1, \cdots, \nu),$$

we have

$$h(f; y, l) = \sum_{j=1}^{\nu} (|\bar{f}^j|^2 - |\bar{f}^{\nu+j}|^2)/2\Im l, \qquad [fg]_l = i \sum_{j=1}^{\nu} (\bar{f}^j \bar{g}^{\nu+j} - \bar{f}^{\nu+j} \bar{g}^j).$$

Now, assuming that $\Im l > 0$ (the case: $\Im l < 0$ can be treated similarly), it is obvious by the above formulae that the linear subspace $\mathfrak{p}$ consisting of all points f of the form $(\bar{f}^1, \cdots, \bar{f}^\nu, 0, \cdots, 0)$, $\bar{f}^j$ being arbitrary complex numbers, is a $(\nu - 1)$-dimensional null plane contained in $[\mathfrak{P}^+(y, l)]$.

Proof of Theorem 2.1. In order to prove the existence of a $(\nu - 1)$-dimensional null plane $\mathfrak{p}$ contained in $\mathfrak{k}_a(l)$, we denote $(\nu - 1)$-dimensional null planes generally by $\mathfrak{p}$ and the set consisting of all $(\nu - 1)$-dimensional null planes by $\mathfrak{N}$. Evidently $\mathfrak{N}$ constitutes an algebraic variety without singularity in the $[C_\nu{}^n - 1]$-dimensional projective space. $\mathfrak{N}$ is therefore compact in the natural topology. Putting $\mathfrak{N}^+(x) = \{\mathfrak{p}; \mathfrak{p} \subset [\mathfrak{P}^+(x, l)]\}$, we have

$$\{\mathfrak{p}; \mathfrak{p} \subset \mathfrak{k}_a(l)\} = \bigcap_x \mathfrak{N}^+(x), \tag{2.12}$$

since $\mathfrak{k}_a(l)$ is the common part of all $[\mathfrak{P}^+(x, l)]$, $a < x < b$. On the other hand, $\mathfrak{N}^+(x)$ is obviously a *compact* subset of $\mathfrak{N}$ and, as Lemma 2.1 shows, *not empty*; furthermore $\mathfrak{N}^+(x)$ *decreases monotonously when x decreases*, since $\mathfrak{P}^+(x, l)$ decreases monotonously. Hence *the intersection* $\bigcap\limits_x \mathfrak{N}^+(x)$ *is not empty*. This shows, combined with (2.12), that there exists at least one null plane $\mathfrak{p}$ contained in $\mathfrak{k}_a(l)$. As to $\mathfrak{k}_b(l)$ the proof will be accomplished in the same way.

From Theorem 2.1 thus proved we can deduce some important conclusions concerning the main spaces $\mathfrak{m}_a(l)$, $\mathfrak{m}_b(l)$. Let $\mathfrak{p}_a$, $\mathfrak{p}_b$ be $(\nu-1)$-dimensional null planes with respect to l contained respectively in $\mathfrak{k}_a(l)$, $\mathfrak{k}_b(l)$. Then, since $\mathfrak{k}_a(l) \subset \mathfrak{m}_a(l)$, $\mathfrak{k}_b(l) \subset \mathfrak{m}_b(l)$, we get first $\dim \mathfrak{m}_a(l) \geqq \nu - 1$, $\dim \mathfrak{m}_b(l) \geqq \nu - 1$. Furthermore $\dim \mathfrak{m}_a(l)$ *and* $\dim \mathfrak{m}_b(l)$ *do not depend on the parameter l, provided* $\Im l \neq 0$, as will be proved in **3** below (see Theorem 3.2). Putting

$$\tau_a = \dim \mathfrak{m}_a(l) - \nu + 1, \qquad \tau_b = \dim \mathfrak{m}_b(l) - \nu + 1 \quad (\Im l \neq 0), \tag{2.13}$$

we obtain therefore two non-negative integers τ_a, τ_b which are characteristic for L.

DEFINITION 2.3. *L will be called a differential operator of the type* (τ_a, τ_b), *τ_a, τ_b being the non-negative integers defined by* (2.13). *The sum $\tau = \tau_a + \tau_b$ will be called the excess index of L.*

Possible values of τ_a, τ_b are obviously $0, 1, 2, \cdots, \nu$; thus *formally self-adjoint differential operators L of the order $n = 2\nu$ can be classified into $(\nu+1)^2$ different types.* This classification is a generalization of H. Weyl's one[10] concerning the differential operators of the second order.

For a later purpose we introduce the space $\mathfrak{m}(l) = \mathfrak{m}_a(l) \cap \mathfrak{m}_b(l)$. Since $\mathfrak{k}_a(l)$ and $\mathfrak{k}_b(l)$ have no common point, $\mathfrak{p}_a$ and $\mathfrak{p}_b$ have also no common point, and therefore $\mathfrak{m}_a(l) + \mathfrak{m}_b(l) \supseteqq \mathfrak{p}_a + \mathfrak{p}_b = \mathfrak{P}$.[11] Hence we get

$$\dim \mathfrak{m}(l) = \tau - 1. \tag{2.14}$$

The space $\mathfrak{m}(l)$ is clearly the set of all points (f) such that $w(x, l, f)$ is square summable in (a, b). *The excess index τ is therefore equal to the number of linearly independent solutions of the differential equation $L[u] = l \cdot u$ which are square summable in (a, b), provided* $\Im l \neq 0$.

3. Green's function. Let $\mathfrak{H}$ be the Hilbert space consisting of all square summable functions in the interval (a, b); the inner product of the functions u, v in $\mathfrak{H}$ will be denoted by (u, v), the norm of u by $\| u \|$.

---

[10] Weyl, [11], pp. 221-231; [12], p. 443; [13].

[11] " + " means the " join " in the sense of the projective geometry.

DEFINITION 3.1. *We denote by* $\mathfrak{D}$ *the linear subspace of* $\mathfrak{H}$ *constituted of all functions* $u(x)$ *satisfying the following four conditions*:

i) $u(x)$ *is square summable in* (a, b),

ii) *in the open interval* (a, b), $u(x)$ *admits continuous derivatives up to the order* $n-1$,

iii) $(n-1)$*-th derivative* $u^{(n-1)}(x)$ *is absolutely continuous in every closed interval* $[x_1, x_2]$, $a < x_1 < x_2 < b$ *(so that* $L[u]$ *can be defined for almost all* x, $a < x < b$*)*,

iv) $L[u]$ *is square summable in* (a, b).

Considered as a linear operator having $\mathfrak{D}$ as its domain, the differential operator L will be denoted by T, i.e. *we put*

$$Tu = L[u], \qquad \text{for } u \in \mathfrak{D}. \tag{3.1}$$

Obviously the bilinear form $[uv](x)$ can be defined for two arbitrary functions u, v in $\mathfrak{D}$ so far as $a < x < b$. Furthermore we infer from Green's formula that, *for arbitrary* u, $v \in \mathfrak{D}$, *the limits* $[uv](a) = \lim_{x \to a} [uv](x)$, $[uv](b) = \lim_{x \to b} [uv](x)$ exist.

Now we shall introduce *Green's function*. Let l be a fixed complex number with $\Im l \neq 0$ and $\mathfrak{p}_a$, $\mathfrak{p}_b$ be two $(\nu-1)$-dimensional null planes with respect to l contained respectively in $\mathfrak{k}_a(l)$, $\mathfrak{k}_b(l)$. Then, choosing n linearly independent vectors $f_1, f_2, \cdots, f_\nu, f_{\nu+1}, \cdots, f_n$ such that $f_\alpha \in \mathfrak{p}_a$ (for $\alpha = 1, 2, \cdots, \nu$), $f_\beta \in \mathfrak{p}_b$ (for $\beta = \nu+1, \cdots, n$) arbitrarily, we put

$$\begin{aligned} G(x, y; l, \mathfrak{p}_a, \mathfrak{p}_b) &= G(y, x; l, \mathfrak{p}_a, \mathfrak{p}_b) \\ &= \sum_{\alpha=1}^{\nu} \sum_{\beta=\nu+1}^{n} F_{\alpha\beta}(l) w_\beta(x) w_\alpha(y), \qquad (x \geqq y), \end{aligned} \tag{3.2}$$

where $w_\gamma(x) = w(x, l, f_\gamma)$ $(\gamma = 1, \cdots, \nu, \nu+1, \cdots, n)$ and $(F_{\alpha\beta}(l))$ means the inverse matrix of $([f_\alpha f_\beta]_l)$. The function $G(x, y; l, \mathfrak{p}_a, \mathfrak{p}_b)$ thus defined will be called *Green's function*. The matrix $([f_\alpha f_\beta]_l)$ is skew-symmetric and, since $f_1, \cdots, f_\nu$ or $f_{\nu+1}, \cdots, f_n$ lie on the null plane $\mathfrak{p}_a$ or $\mathfrak{p}_b$, respectively, we have

$$[f_\alpha f_\beta]_l = 0, \qquad \text{for } \alpha \leqq \nu, \beta \leqq \nu, \text{ and for } \alpha > \nu, \beta > \nu. \tag{3.3}$$

Consequently we see

$$F_{\alpha\beta}(l) = -F_{\beta\alpha}(l), \tag{3.4}$$

$$(3.5) \qquad F_{\alpha\beta}(l)=0, \qquad \text{for } \alpha \leqq \nu, \beta \leqq \nu, \text{ and for } \alpha > \nu, \beta > \nu.$$

Introduce now the important matrix

$$(3.6) \qquad M^{jk}(l, \mathfrak{p}_a, \mathfrak{p}_b) = \sum_{\alpha=1}^{\nu} \sum_{\beta=\nu+1}^{n} F_{\alpha\beta}(l) f_\beta{}^j f_\alpha{}^k.$$

Then we obtain, using (3.3) and (3.5),

$$\sum_k \sum_m M^{mj}(l, \mathfrak{p}_a, \mathfrak{p}_b)[s_m(l)s_k(l)] \cdot f_\gamma{}^k = \begin{cases} f_\gamma{}^j, & (1 \leqq \gamma \leqq \nu), \\ 0, & (\nu < \gamma \leqq n), \end{cases}$$

proving the relation

$$(3.7) \qquad \sum_k \sum_m M^{mj}(l, \mathfrak{p}_a, \mathfrak{p}_b)[s_m(l)s_k(l)] \cdot f^k = \begin{cases} f^j, & (f \in \mathfrak{p}_a), \\ 0, & (f \in \mathfrak{p}_b). \end{cases}$$

Similarly we get

$$(3.8) \qquad \sum_k \sum_m M^{jm}(l, \mathfrak{p}_a, \mathfrak{p}_b)[s_m(l)s_k(l)] \cdot f^k = \begin{cases} 0, & (f \in \mathfrak{p}_a), \\ -f^j, & (f \in \mathfrak{p}_b). \end{cases}$$

These relations show that $M^{jk}(l, \mathfrak{p}_a, \mathfrak{p}_b)$ *is determined uniquely by* l, $\mathfrak{p}_a, \mathfrak{p}_b$ *and does not depend on the choice of the points* $f_1, \cdots, f_\nu$ *on* $\mathfrak{p}_a$ *and* $f_{\nu+1}, \cdots, f_n$ *on* $\mathfrak{p}_b$. By using the matrix $M^{jk}(l, \mathfrak{p}_a, \mathfrak{p}_b)$, Green's function can be rewritten as

$$(3.9) \qquad G(x, y; l, \mathfrak{p}_a, \mathfrak{p}_b) = \sum M^{jk}(l, \mathfrak{p}_a, \mathfrak{p}_b) s_j(x, l) s_k(y, l), \qquad (x \geqq y).$$

Hence Green's function is also determined uniquely by l, $\mathfrak{p}_a$, $\mathfrak{p}_b$ *and does not depend on the choice of* $f_1, \cdots, f_\nu, \cdots, f_n$ *appearing in its definition.* Incidentally, we infer from (3.7) and (3.8) the relation

$$(3.10) \qquad \sum_m \{M^{mj}(l, \mathfrak{p}_a, \mathfrak{p}_b) - M^{jm}(l, \mathfrak{p}_a, \mathfrak{p}_b)\}[s_m(l)s_k(l)] = \delta_k{}^j.$$

DEFINITION 3.2. $G(l, \mathfrak{p}_a, \mathfrak{p}_b)$ *will denote the integral operator with the kernel* $G(x, y; l, \mathfrak{p}_a, \mathfrak{p}_b)$, *that is the operator defined by*

$$(3.11) \qquad \begin{aligned} G(l, \mathfrak{p}_a, \mathfrak{p}_b) v(x) &= \int_a^b G(x, y; l, \mathfrak{p}_a, \mathfrak{p}_b) v(y)\, dy \\ &= \sum_{\alpha=1}^{\nu} \sum_{\beta=\nu+1}^{n} F_{\alpha\beta}(l) \{ w_\beta(x) \int_a^x w_\alpha(y) v(y)\, dy + w_\alpha(x) \int_x^b w_\beta(y) v(y)\, dy \}, \end{aligned}$$

$v(x)$ *being an arbitrary function in* $\mathfrak{H}$.

Since $f_\alpha \in \mathfrak{p}_a \subset \mathfrak{k}_a(l) \subset \mathfrak{m}_a(l)$ $(\alpha = 1, \cdots, \nu)$, the functions $w_\alpha(x) = w(x, l, f_\alpha)$ $(\alpha = 1, 2, \cdots, \nu)$ are square summable in $(a, x]$, and similarly the functions $w_\beta(x)$ $(\beta = \nu + 1, \cdots, n)$ are square summable in $[x, b)$. Hence *the integrals in* (3.11) *converge absolutely*; thus the function $u(x) = G(l, \mathfrak{p}_a, \mathfrak{p}_b) v(x)$ is defined for all $v(x)$ in $\mathfrak{H}$. Now we *shall prove* *that the function* $u(x)$ *thus defined admits continuous derivatives up to the order* $n-1$, *that the* $(n-1)$*-th derivative* $u^{(n-1)}(x)$ *is absolutely continuous in every closed interval* $[x_1, x_2]$, $a < x_1 < x_2 < b$, *and that* $u(x)$ *satisfies the differential equation*:

$$L[u] - l \cdot u = v. \tag{3.12}$$

From the relation

$$[f_\alpha f_\beta]_l = [w_\alpha w_\beta] = \sum_{j+k \leqq n-1} B_{jk}(x) w_\alpha{}^{(j)}(x) w_\beta{}^{(k)}(x)$$

we get readily

$$\sum_{\alpha=1}^{n} \sum_{\beta=1}^{n} F_{\alpha\beta}(l) w_\alpha{}^{(j)}(x) w_\beta{}^{(k)}(x) = B^{-1}{}_{jk}(x), \tag{3.13}$$

where $(B^{-1}{}_{jk})$ means the inverse matrix of (B_{jk}). Since

$$B_{jk}(x) = (-1)^{j+1} p_0(x) \quad (\text{for } j+k = n-1), = 0 \quad (\text{for } j+k > n-1), \tag{3.14}$$

we have

$$B^{-1}{}_{jk}(x) = (-1)^{j}/p_0(x) \quad (\text{for } j+k = n-1), = 0 \quad (\text{for } j+k < n-1).$$

Combined with (3.4) and (3.5), (3.13) yields therefore

$$\sum_{\alpha=1}^{\nu} \sum_{\beta=\nu+1}^{n} F_{\alpha\beta}(l)\{w_\beta{}^{(m)}(x) w_\alpha(x) - w_\alpha{}^{(m)}(x) w_\beta(x)\} = 0, \quad (m = 0, 1, \cdots, n-2), \tag{3.15}$$

$$\sum_{\alpha=1}^{\nu} \sum_{\beta=\nu+1}^{n} F_{\alpha\beta}(l)\{w_\beta{}^{(n-1)}(x) w_\alpha(x) - w_\alpha{}^{(n-1)}(x) w_\beta(x)\} = 1/p_0(x). \tag{3.16}$$

Using (3.15), we infer readily that $u(x)$ is differentiable in (a, b) up to the order $n-1$ and that the m-th derivatives $u^{(m)}(x)$ are given by

$$u^{(m)}(x) = \sum_{\alpha=1}^{\nu} \sum_{\beta=\nu+1}^{n} F_{\alpha\beta}(l)\{w_\beta{}^{(m)}(x) \int_a^x w_\alpha v\, dy + w_\alpha{}^{(m)}(x) \int_x^b w_\beta v\, dy\}, \quad (m = 1, 2, \cdots, n-1). \tag{3.17}$$

Evidently $u^{(m)}(x)$ $(m = 0, \cdots, n-1)$ are absolutely continuous in every closed interval $[x_1, x_2]$, $a < x_1 < x_2 < b$. Especially $u^{(n-1)}(x)$ is differentiable almost everywhere in (a, b) and, as one readily infers by (3.16), the derivative $u^{(n)}(x)$ is given by

$$u^{(n)}(x) = v(x)/p_0(x) + \sum_{\alpha=1}^{\nu} \sum_{\beta=\nu+1}^{n} F_{\alpha\beta}(l)\{w_\beta{}^{(n)}(x) \int_a^x w_\alpha v\, dy + w_\alpha{}^{(n)}(x) \int_x^b w_\beta v\, dy\}.$$

Combined with (3.17), this shows clearly that $u(x)$ satisfies the differential equation (3.12).

Incidentally, from (3.14), (3.15) and (3.16) follows the identity

$$(3.18) \qquad \sum_{\alpha=1}^{\nu} \sum_{\beta=\nu+1}^{n} F_{\alpha\beta}(l)\{[w_\beta v](x) w_\alpha(x) - [w_\alpha v](x) w_\beta(x)\} = v(x),$$

which will be used in **5** below.

THEOREM 3.1. *$G(l, \mathfrak{p}_a, \mathfrak{p}_b)$ is a bounded linear operator whose norm does not exceed $|\Im l|^{-1}$, provided $\Im l \neq 0$. Furthermore, for an arbitrary function $v(x)$ in $\mathfrak{H}$, the function $u(x) = G(l, \mathfrak{p}_a, \mathfrak{p}_b) v(x)$ belongs to $\mathfrak{D}$ and satisfies the differential equation $L[u] - l \cdot u = v$.*

Proof.[12] To show that $G(l, \mathfrak{p}_a, \mathfrak{p}_b)$ is a bounded operator, we take two points t, z such that $a < t < z < b$ arbitrarily and put

$$u_1(x) = \sum_{\alpha=1}^{\nu} \sum_{\beta=\nu+1}^{n} F_{\alpha\beta}(l)\{w_\beta(x) \int_t^x w_\alpha v \, dy + w_\alpha(x) \int_x^z w_\beta v \, dy\}.$$

Then $u_1(x)$ satisfies also the differential equation $L[u_1] - l \cdot u_1 = v$. Hence we get, using Green's formula,

$$(3.19) \qquad \int_t^z |u_1|^2 dx + (1/2i\Im l) \int_t^z (v\bar{u}_1 - u_1 \bar{v}) dx \\ = \{[u_1\bar{u}_1](z) - [u_1\bar{u}_1](t)\}/2i\Im l.$$

Now, as one readily infers by (3.15), $u_1^{(m)}(z)$ $(m = 0, 1, \cdots, n-1)$ are given by $u_1^{(m)}(z) = w^{(m)}(z, l, g)$, where

$$g = \sum_{\beta=\nu+1}^{n} \{\sum_{\alpha=1}^{\nu} F_{\alpha\beta}(l) \cdot \int_t^z w_\alpha v \, dy\} \cdot f_\beta.$$

We have therefore

$$[u_1\bar{u}_1](z)/2i\Im l = h(g; z, l).$$

On the other hand, since f_β $(\beta = \nu + 1, \cdots, n)$ lie on $\mathfrak{p}_b$, g lies also in $\mathfrak{p}_b$, while we have $\mathfrak{p}_b \subset \mathfrak{k}_b(l) \subset \mathfrak{P}^-(z, l)$. Hence $h(g; z, l)$ is negative and therefore $[u_1\bar{u}_1](z)/2i\Im l < 0$; similarly we get $[u_1\bar{u}_1](t)/2i\Im l > 0$. Using these inequalities, we conclude from (3.19) the inequality

$$\int_t^z |u_1|^2 dx \leqq |\Im l|^{-2} \int_t^z |v|^2 dx.$$

Now, making here $t \to a$, $z \to b$, we infer from this the inequality

$$(3.20) \qquad \| u \| \leqq |\Im l|^{-1} \| v \|, \qquad (u = G(l, \mathfrak{p}_a, \mathfrak{p}_b) v),$$

---

[12] We follow the method of H. Weyl. See Weyl, [11], pp. 228-231.

proving that $G(l, \mathfrak{p}_a, \mathfrak{p}_b)$ is a bounded operator and that its norm satisfies the inequality

$$|||\, G(l, \mathfrak{p}_a, \mathfrak{p}_b)\, ||| \leqq |\, \Im l\, |^{-1}. \tag{3.21}$$

As was already proved, the function $u(x)$ satisfies the differential equation (3.12) and the conditions ii) and iii) in Definition 3.1, while (3.20) shows that $u(x)$ satisfies the condition i) in Definition 3.1. Now, using these results, we get

$$\| L[u] \| = \| l \cdot u + v \| \leqq (1 + |\, l\, | / |\, \Im l\, |) \cdot \| v \|,$$

proving that u satisfies also the condition iv) in Definition 3.1. Thus we see that the function $u(x)$ belongs to $\mathfrak{D}$.

COROLLARY. *We have*

$$(T - l)\, G(l, \mathfrak{p}_a, \mathfrak{p}_b) = 1, \qquad (\Im l \neq 0). \tag{3.22}$$

By virtue of Theorem 3.1, we can now prove the following theorem cited already in **2** above:

THEOREM 3.2. Dim $\mathfrak{m}_a(l)$ *and* dim $\mathfrak{m}_b(l)$ *do not depend on* l, *provided* $\Im l \neq 0$.

To prove this, we first introduce

DEFINITION 3.3. *For arbitrary l with $\Im l \neq 0$, we denote by $\mathfrak{E}(l)$ the eigenspace of the operator T corresponding to the eigenvalue l.*

Evidently $\mathfrak{E}(l)$ can be written as

$$\mathfrak{E}(l) = \{ w(x, l, f)\, ;\ (f) \,\varepsilon\, \mathfrak{m}(l) \text{ or } f = 0 \}. \tag{3.23}$$

Combined with (2.14), this yields

$$\dim \mathfrak{E}(l) = \tau, \tag{3.24}$$

τ being the excess index of L. We have now to prove that τ is independent on l.[13]

Choose an arbitrary complex number l^0 with $\Im l^0 \neq 0$, put, for simplicity's sake, $G^0 = G(l^0, \mathfrak{p}_a{}^0, \mathfrak{p}_b{}^0)$, $\mathfrak{p}_a{}^0$, $\mathfrak{p}_b{}^0$ being $(\nu - 1)$-dimensional null planes with respect to l^0 contained respectively in $\mathfrak{k}_a(l^0)$, $\mathfrak{k}_b(l^0)$, and introduce the operator $K(l) = 1 - (l - l^0)\, G^0$. Again we denote by $C(l^0)$ the circular region in the l-plane consisting of all l satisfying $|\, l - l^0\, | < |\, \Im l^0\, |$. Then we infer by (3.21) that, for arbitrary l in $C(l^0)$, the operator $K(l)$ has the inverse $K(l)^{-1}$. Indeed the inverse is given by the power series

---

[13] We follow the method of H. Weyl. See Weyl, [13], pp. 238-247. Cf. also Stone, [7], Theorem 9.8.

$$K(l)^{-1} = 1 + \sum_{m=1}^{\infty} (l - l^0)^m (G^0)^m, \tag{3.25}$$

converging for $l \varepsilon C(l^0)$ in the sense of the norm: $|||\ \ |||$. Thus in case $l \varepsilon C(l^0)$, $K(l)$ is a one-to-one transformation mapping $\mathfrak{H}$ on itself. Now, using (3.22), we get readily $(T - l^0) \cdot K(l) u = (T - l) \cdot u$, (for all $u \varepsilon \mathfrak{H}$). This shows that $K(l)$ *maps* $\mathfrak{E}(l)$ *on* $\mathfrak{E}(l^0)$ *one-to-one.* Hence $\dim \mathfrak{E}(l)$ is constant in $C(l^0)$, while, by choosing l^0 suitably, an arbitrary compact domain D in the half plane $\Im l > 0$ (or $\Im l < 0$) can be covered by $C(l^0)$. Consequently $\dim \mathfrak{E}(l)$ is constant in each half plane $\Im l > 0$, $\Im l < 0$. On the other hand, from (2.6) and (3.23) follows $\mathfrak{E}(\bar{l}) = \bar{\mathfrak{E}}(l)$. Hence $\dim \mathfrak{E}(l)$ *does not depend on* l, provided $\Im l \neq 0$.

To deduce Theorem 3.2 from this result, we choose a fixed point c, $a < c < b$, arbitrarily, and *observe the operator* L *in the interval* $(a, c]$ *instead of* (a, b). *The meanings of the notations* $\mathfrak{H}$, $\mathfrak{m}(l)$, $\mathfrak{E}(l)$, *etc. are to be changed correspondingly, e. g.* $\mathfrak{H}$ *denotes the Hilbert space consisting of all square summable functions defined in* $(a, c]$, $\mathfrak{m}(l)$ *the subspace of* $\mathfrak{P}$ *consisting of all* f *such that* $w(x, l, f)$ *is square summable in* $(a, c]$, etc. Then, since every solution $w(x, l, f)$ is continuous in c, we have $\mathfrak{m}_c(l) = \mathfrak{P}$ and therefore $\mathfrak{m}(l) = \mathfrak{m}_a(l)$, while, as was proved above, $\dim \mathfrak{m}(l) = \dim \mathfrak{E}(l) - 1$ does not depend on l. Hence $\dim \mathfrak{m}_a(l)$ is independent on l. It can be proved similarly that $\dim \mathfrak{m}_b(l)$ does not depend on l.

In this connection, we shall prove furthermore

THEOREM 3.3. *As functions of* l, *the main spaces* $\mathfrak{m}_a(l)$, $\mathfrak{m}_b(l)$, *and their intersection* $\mathfrak{m}(l)$ *are holomorphic, provided* $\Im l \neq 0$, *in the sense that, for every* l_0 *with* $\Im l_0 \neq 0$, *the suitably normalized*[14] *Plücker coordinates* $\mathfrak{m}_a^{ij\ldots k}(l), \mathfrak{m}_b^{ij\ldots k}(l), \mathfrak{m}^{ij\ldots k}(l)$ *of* $\mathfrak{m}_a(l)$, $\mathfrak{m}_b(l)$, $\mathfrak{m}(l)$ *are holomorphic with respect to* l *in some neighborhood of* l_0.

To prove this,[15] we introduce

DEFINITION 3.4. *Let* $r(x, l)$ *be a square summable function of* x *defined in* (a, b) *depending on the parameter* l *and* D *be a domain in the* l*-plane.*

---

[14] Generally, the homogeneous coordinates of a geometric object are said to be *normalized,* if one of the coordinates is equal to 1.

[15] We follow the method of H. Weyl. See Weyl, [13], pp. 238-247. I am indebted to Prof. Weyl for suggesting to me to apply here his direct method.

Then $r(x,l)$ *is said to be holomorphic with respect to* l *in* D *in the sense of the norm:* $\| \quad \|$, *if, for every point* l_0 *in* D, $r(x,l)$ *can be expanded into the power series*

$$r(x,l) = \sum_{m=0}^{\infty} r_m(x)(l-l_0)^m, \qquad (r_m(x) \in \mathfrak{H})$$

converging in the sense of the norm: $\| \quad \|$ *in some neighborhood of* l_0.

LEMMA 3.1. *Let a compact domain* D *in the half-plane* $\Im l > 0$ (*or* $\Im l < 0$) *be given. Then we can choose a base* $\{r_\sigma(x,l); \sigma = 1,2,\cdots,\tau\}$ *of the functional space* $\mathfrak{E}(l)$ *so that* $r_\sigma(x,l)$ *are holomorphic in* D *with respect to* l *in the sense of the norm:* $\| \quad \|$ *and that, for every fixed* x, $a < x < b$, *the derivatives* $d^m r_\sigma(x,l)/dx^m$ $(m = 0,1,\cdots,n-1)$ *are holomorphic in* D *with respect to* l (*in the usual sense*).

Proof. As was proved above, we have $\mathfrak{E}(l) = K(l)^{-1}\mathfrak{E}(l^0)$ for $l \in C(l^0)$. Hence, choosing a base $\{r^0{}_\sigma(x); \sigma = 1,\cdots,\tau\}$ of the space $\mathfrak{E}(l^0)$ arbitrarily and putting

$$r_\sigma(x,l) = K(l)^{-1} r^0{}_\sigma(x), \tag{3.26}$$

we obtain, for arbitrary $l \in C(l^0)$, the base $\{r_\sigma(x,l); \sigma = 1,\cdots,\tau\}$ of the space $\mathfrak{E}(l)$. Now it follows from (3.25) that $r_\sigma(x,l)$ are holomorphic in $C(l^0)$ with respect to l in the sense of the norm: $\| \quad \|$. Again we get from (3.26) $r_\sigma(x,l) = r^0{}_\sigma(x) + (l-l^0)G^0 \cdot r_\sigma(x,l)$, which yields, combined with (3.17),

$$\begin{aligned} d^m r_\sigma(x,l)/dx^m = {} & d^m r^0{}_\sigma(x)/dx^m + (l-l^0) \sum_{\alpha=1}^{\nu} \sum_{\beta=\nu+1}^{n} F_{\alpha\beta}(l^0) \\ & \cdot \{(d^m w^0{}_\beta/dx^m) \cdot \int_a^x w^0{}_\alpha(y) r_\sigma(y,l)\,dy + (d^m w^0{}_\alpha/dx^m) \\ & \cdot \int_x^b w^0{}_\beta(y) r_\sigma(y,l)\,dy\}, \qquad (m = 0,1,\cdots,n-1). \end{aligned}$$

Whence we infer that $d^m r_\sigma(x,l)/dx^m$ $(m = 0,1,\cdots,n-1)$ are also holomorphic with respect to l in $C(l^0)$. This proves Lemma 3.1, since, by a suitable choice of l^0, D can be covered by $C(l^0)$.

Proof of Theorem 3.3. Let D be an arbitrary compact domain in the half plane $\Im l > 0$ (or $\Im l < 0$) and $\{r_\sigma(x,l); \sigma = 1,2,\cdots,\tau\}$ be the base of $\mathfrak{E}(l)$ mentioned in Lemma 3.1 above. Then, putting

$$r_\sigma(x,l) = \sum_{j=1}^{n} s_j(x,l) \cdot f_\sigma{}^j(l), \qquad (\sigma = 1,2,\cdots,\tau),$$

we obtain τ linearly independent points $f_\sigma(l)$ $(\sigma = 1,2,\cdots,\tau)$ generating the linear space $\mathfrak{m}(l)$. The coordinates $f_\sigma{}^j(l)$ are determined by the system of linear equations

$$\sum_{j=1}^{n} s_j{}^{(m)}(x, l)\cdot f_\sigma{}^j(l) = r_\sigma{}^{(m)}(x, l), \qquad (m = 0, 1, \cdots, n-1)$$

for a fixed point x, while $s_j{}^{(m)}(x, l)$, $r_\sigma{}^{(m)}(x, l)$ are holomorphic with respect to l in D and $[s_1 s_2 \cdots s_n] = 1$. Hence $f_\sigma{}^j(l)$ are also holomorphic in D, proving that the linear space $\mathfrak{m}(l)$ depends on l holomorphically, provided $\Im l \neq 0$.

Consider now the operator L in the interval $(a, c]$, $a < c < b$, instead of (a, b), as in the proof of Theorem 3.2. Then $\mathfrak{m}_a(l)$ coincides with $\mathfrak{m}(l)$. Hence $\mathfrak{m}_a(l)$ is holomorphic with respect to l, and similarly $\mathfrak{m}_b(l)$ depends also holomorphically on l, provided $\Im l \neq 0$.

Now we turn to investigate the operator T. First we prove

THEOREM 3.4. *T is a closed operator. The domain $\mathfrak{D}$ of T consists of all elements $u \varepsilon \mathfrak{H}$ having the following form:*

$$(3.27) \qquad u = G(l, \mathfrak{p}_a, \mathfrak{p}_b)v + w, \qquad (v \varepsilon \mathfrak{H}, w \varepsilon \mathfrak{E}(l)),$$

l being a fixed complex number with $\Im l \neq 0$.[16]

Proof. By virtue of Theorem 3.1, it is obvious that every u of the form (3.27) belongs to $\mathfrak{D}$. Conversely, let an arbitrary element $u \varepsilon \mathfrak{D}$ be given. Then, putting

$$v = (T - l)u, \qquad w = u - G(l, \mathfrak{p}_a, \mathfrak{p}_b)v,$$

we infer, by (3.22), $(T - l)w = 0$, proving that w belongs to $\mathfrak{E}(l)$. Thus we see that u has the form (3.27). To prove that T is a closed operator, consider a sequence $\{u_m\}$, $u_m \varepsilon \mathfrak{D}$, such that the limits $u = \lim u_m$, $u^* = \lim Tu_m$, exist. As was proved above, u_m can be written as

$$u_m = G(l, \mathfrak{p}_a, \mathfrak{p}_b)(Tu_m - lu_m) + w_m, \qquad (w_m \varepsilon \mathfrak{E}(l)).$$

Hence there exists the limit $w = \lim w_m$ and we have

$$(3.28) \qquad u = G(l, \mathfrak{p}_a, \mathfrak{p}_b)(u^* - lu) + w, \qquad (w \varepsilon \mathfrak{E}(l)),$$

proving that u belongs to $\mathfrak{D}$. Again, using (3.22), we get, from (3.28), $(T - l)u = u^* - lu$, or $Tu = u^*$; thus T is a closed operator, q. e. d.

---

[16] The closures and adjoints of ordinary differential operators were determined by I. Halperin under some restrictions imposed on the coefficients of the differential operators. See Halperin, [2]. The operator L under consideration is not necessarily subject to Halperin's restrictions.

It is to be noted here that, *for the element* u *having the form* (3.27) we have $(T-l)u=v$, as one readily infers by (3.22).

It is obvious that the *adjoint operator* $G^*(l, \mathfrak{p}_a, \mathfrak{p}_b)$ of $G(l, \mathfrak{p}_a, \mathfrak{p}_b)$ is given by

$$G^*(l, \mathfrak{p}_a, \mathfrak{p}_b)v(x) = \int_a^b \bar{G}(x, y; l, \mathfrak{p}_a, \mathfrak{p}_b)v(y)dy. \tag{3.29}$$

Now, using this, we shall determine the *adjoint operator of* T, *which will be denoted by* T_0.

THEOREM 3.5. *The domain* $\mathfrak{D}_0$ *of the adjoint operator* T_0 *of* T *consists of all elements* $u \varepsilon \mathfrak{D}$ *having the following form:*

$$u = G(l, \mathfrak{p}_a, \mathfrak{p}_b)v, \qquad v \varepsilon \mathfrak{H} - \bar{\mathfrak{E}}(l)^{17}; \tag{3.30}$$

thus $\mathfrak{D}_0$ *is a subspace of* $\mathfrak{D}$. *The adjoint operator* T_0 *is obtained from* T *by restricting the domain* $\mathfrak{D}$ *of* T *to* $\mathfrak{D}_0$. T_0 *is a symmetric operator.*

Proof. The relation: $u^* = T_0 u$ means, by definition, that

$$(u, Tu_1) = (u^*, u_1) \tag{3.31}$$

holds for every element $u_1 \varepsilon \mathfrak{D}$. Obviously the domain $\mathfrak{D}$ remains unchanged by the conjugation $u(x) \to \overline{u(x)}$. Hence, using (3.29), we infer by Theorem 3.4 that $\mathfrak{D}$ consists of all elements of the form $u_1 = G^*(l, \mathfrak{p}_a, \mathfrak{p}_b)v_1 + \bar{w}$, $(v_1 \varepsilon \mathfrak{H},\ w \varepsilon \mathfrak{E}(l))$. By inserting this expression in (3.31), the condition (3.31) becomes $(u + lGu - Gu^*, v_1) + (lu - u^*, \bar{w}) = 0$, where $G = G(l, \mathfrak{p}_a, \mathfrak{p}_b)$. (3.31) is therefore equivalent to

$$\begin{cases} u = G(l, \mathfrak{p}_a, \mathfrak{p}_b)(u^* - lu), \\ (u^* - lu, \bar{w}) = 0, \qquad \text{for all } w \varepsilon \mathfrak{E}(l). \end{cases} \tag{3.32}$$

Now it is obvious that every u satisfying (3.32) has the form (3.30). Assume conversely u to have the form (3.30). Then putting $u^* = v + lu$, we get immediately (3.32). Thus $\mathfrak{D}_0$ consists of all elements u of the form (3.30). Furthermore, for such u, we get, using (3.22) and (3.32), $Tu = u^* = T_0 u$. Thus T is an extension of T_0, q. e. d.

It is to be noted here that *the deficiency index*[18] *of the symmetric operator* T_0 is (τ, τ).

Finally we shall prove

THEOREM 3.6.[19] *For arbitrary fixed* l *with* $\mathfrak{J}l \neq 0$, *the domain* $\mathfrak{D}$ *is decomposed as a linear space into the direct sum:*

---

[17] $\mathfrak{H} - \mathfrak{E}$ means the orthogonal complement of the subspace $\mathfrak{E}$ in $\mathfrak{H}$.

[18] See Stone, [7], Chap. IX.

[19] This theorem is essentially reduced to Theorem 9.4 in Stone, [7]. Here we shall give another proof based on Green's function.

$$(3.33)\qquad \mathfrak{D} = \mathfrak{D}_0 + \mathfrak{E}(l) + \bar{\mathfrak{E}}(l), \qquad (\Im l \neq 0).$$

Proof. It is sufficient to show that every element $u \varepsilon \mathfrak{D}$ is decomposed uniquely in the form

$$(3.34)\qquad u = u_0 + w_1 + \bar{w}_2, \qquad (u_0 \varepsilon \mathfrak{D}_0, w_1, w_2 \varepsilon \mathfrak{E}(l)).$$

By Theorem 3.4, u is represented as $u = Gv + w_3$ $(w_3 \varepsilon \mathfrak{E}(l))$, where G means $G(l, \mathfrak{p}_a, \mathfrak{p}_b)$. Decompose now v into two parts: $v = v_0 + \bar{w}$, $(v_0 \varepsilon H - \bar{\mathfrak{E}}(l), w \varepsilon \mathfrak{E}(l))$ and put $u_0 = Gv_0$, $w_1 = G\bar{w} + \bar{w}/2i\Im l + w_3$, $w_2 = w/2i\Im l$. Then we have the desired decomposition $u = u_0 + w_1 + \bar{w}_2$, $(w_1, w_2 \varepsilon \mathfrak{E}(l))$.

To prove the uniqueness of the decomposition, it is sufficient to show that the relation

$$(3.35)\qquad 0 = u_0 + w_1 + \bar{w}_2, \qquad (u_0 \varepsilon \mathfrak{D}, w_1, w_2 \varepsilon \mathfrak{E}(l))$$

implies $u_0 = w_1 = w_2 = 0$. Assuming (3.35), we have $2i\Im l \cdot (w_1, w_1) = (w_1, T_0 u_0) - (Tw_1, u_0) = 0$, proving that w_1 is equal to 0. Similarly we get $w_2 = 0$, and therefore $u = 0$, q. e. d.

COROLLARY. *We have*

$$(3.36)\qquad \dim (\mathfrak{D}/\mathfrak{D}_0) = 2\tau,$$

τ *being the excess index of* L.

4. Boundary conditions.

First we shall study the properties of the skew-symmetric bilinear functionals $[uv](a)$, $[uv](b)$ defined by

$$(4.1)\qquad [uv](a) = \lim_{x\to a} [uv](x), \qquad [uv](b) = \lim_{x\to b} [uv](x), \quad (u, v \varepsilon \mathfrak{D}).$$

We start with the important formula:

$$(4.2)\qquad (Tu, v) - (u, Tv) = [u\bar{v}](b) - [u\bar{v}](a),$$

deduced immediately from Green's formula. Equation (4.2) shows that the necessary and sufficient condition for u to be contained in $\mathfrak{D}_0$ is

$$(4.3)\qquad [u\bar{v}](b) - [u\bar{v}](a) = 0, \qquad \text{for all } v \varepsilon \mathfrak{D}.$$

On the other hand, an arbitrary function $v \varepsilon \mathfrak{D}$ can be decomposed as

$$(4.4)\qquad v = v_1 + v_2, \qquad (v_1, v_2 \varepsilon \mathfrak{D}),$$

so that

$$\begin{cases} v_2(x) = 0, \quad v_1(x) = v(x), & \text{for } a < x \leqq c, \\ v_1(x) = 0, \quad v_2(x) = v(x), & \text{for } d \leqq x < b, \end{cases}$$

c, d being fixed points such that $a < c < d < b$. For arbitrary $u \varepsilon \mathfrak{D}_0$, we get therefore, by (4. 1) and (4. 3), $[u\bar{v}](b) = [u\bar{v}_2](b) = [u\bar{v}_2](b) - [u\bar{v}_2](a) = 0$, and similarly $[u\bar{v}](a) = 0$. Thus we see, using the invariance of $\mathfrak{D}$ under the conjugation $u \rightarrow \bar{u}$, that *an element u of $\mathfrak{D}$ lies in $\mathfrak{D}_0$ if and only if*

$$(4.5) \qquad [uv](a) = 0, \qquad [uv](b) = 0, \qquad \text{for all } v \varepsilon \mathfrak{D}.$$

Denote by $\mathfrak{D}_a$ (or $\mathfrak{D}_b$) the subspace of $\mathfrak{D}$ consisting of all u satisfying the boundary conditions $[uv](a) = 0$ (or $[uv](b) = 0$) for all $v \varepsilon \mathfrak{D}$. Then we infer by (4. 5)

$$(4.6) \qquad \mathfrak{D}_a \cap \mathfrak{D}_b = \mathfrak{D}_0,$$

while, since v_1, v_2 in (4. 4) are contained in $\mathfrak{D}_b$, $\mathfrak{D}_a$, respectively, we have

$$(4.7) \qquad \mathfrak{D}_a + \mathfrak{D}_b = \mathfrak{D}.$$

Using (3. 36), we conclude from (4. 6) and (4. 7)

$$(4.8) \qquad \dim(\mathfrak{D}/\mathfrak{D}_a) + \dim(\mathfrak{D}/\mathfrak{D}_b) = 2\tau.$$

Moreover *we have*

$$(4.9) \qquad \dim(\mathfrak{D}/\mathfrak{D}_a) = 2\tau_a, \qquad \dim(\mathfrak{D}/\mathfrak{D}_b) = 2\tau_b.$$

To prove this, we observe the operator L in the interval $(a, c]$, $a < c < b$, instead of (a, b), as in the proof of Theorem 3. 2. Then we have $\tau_c = \nu$, and therefore

$$(4.10) \qquad \tau = \tau_a + \tau_c = \tau_a + \nu.$$

On the other hand, we have

$$\mathfrak{D}_c = \{u; u \varepsilon \mathfrak{D}, u(c) = u'(c) = \cdots = u^{(n-1)}(c) = 0\},$$

and consequently $\dim(\mathfrak{D}/\mathfrak{D}_c) = 2\nu$. Combined with (4. 8) and (4. 10), this proves the first formula in (4. 9). The second formula can be proved similarly.

Evidently $[uv](a)$, $[uv](b)$ *can be considered as non-singular bilinear functionals defined respectively in the residual spaces $\mathfrak{D}/\mathfrak{D}_a$, $\mathfrak{D}/\mathfrak{D}_b$.* From (4. 1) it follows that the functionals $[uv](a)$, $[uv](b)$ *are real* in the sense that

$$(4.11) \qquad [\bar{u}\bar{v}](a) = \overline{[uv](a)}, \qquad [\bar{u}\bar{v}](b) = \overline{[uv](b)}.$$

Combined with the invariance of $\mathfrak{D}$ under the conjugation $u \rightarrow \bar{u}$, this yields furthermore

$$(4.12) \qquad \bar{\mathfrak{D}}_a = \mathfrak{D}_a, \qquad \bar{\mathfrak{D}}_b = \mathfrak{D}_b, \qquad \bar{\mathfrak{D}}_0 = \mathfrak{D}_0.$$

By a boundary condition at a, we shall mean a condition for $u \varepsilon \mathfrak{D}$ of the form $[\phi u](a) = 0$, where ϕ means a fixed function belonging to $\mathfrak{D}$. The boundary condition will be called *real,* if ϕ is a real-valued function. A finite number of boundary conditions

$$[\phi_1 u](a) = 0, [\phi_2 u](a) = 0, \cdots, [\phi_m u](a) = 0$$

will be called *linearly independent,* if every linear combination $\Sigma \gamma^j [\phi_j u](a)$ does not vanish identically in u except for the case: $\gamma^1 = \gamma^2 = \cdots = \gamma^m = 0$. *The linear independence of the boundary conditions $[\phi_1 u](a) = 0, \cdots, [\phi_m u](a) = 0$ is therefore equivalent to the linear independence of the functions $\phi_1, \phi_2, \cdots, \phi_m$* $\operatorname{mod} \mathfrak{D}_a$. Boundary conditions at b, their reality and their linear independence are to be defined similarly.

DEFINITION 4. 1. *A system of m linearly independent boundary conditions $[\phi_j u](a) = 0$ (or $[\phi_j u](b) = 0$) $(j = 1, 2, \cdots, m)$ will be called self-adjoint, if ϕ_j satisfy $[\phi_j \phi_k](a) = 0$ (or $[\phi_j \phi_k](b) = 0$) $(j, k = 1, 2, \cdots, m)$ and $m = \tau_a$ (or $m = \tau_b$).*

Our main purpose is to investigate the "eigenvalue problem" for the differential equation $L[u] = \lambda u$ under the self-adjoint systems of real boundary conditions given at the both ends of the interval a and b. Let

$$(4.13)_a \qquad [\phi_{aj} u](a) = 0, \qquad (j = 1, 2, \cdots, \tau_a),$$

$$(4.13)_b \qquad [\phi_{bj} u](b) = 0, \qquad (j = 1, 2, \cdots, \tau_b),$$

be the given self-adjoint systems of the linearly independent real boundary conditions. Then we denote by $\mathfrak{D}_\phi$ the subspace of $\mathfrak{D}$ consisting of all u satisfying the conditions $(4.13)_a$, $(4.13)_b$ and by H the operator with the domain $\mathfrak{D}_\phi$ mapping every $u \varepsilon \mathfrak{D}_\phi$ in $L[u]$, i. e. we put $Hu = L[u]$, for $u \varepsilon \mathfrak{D}_\phi$. Obviously we have $\mathfrak{D}_0 \subseteqq \mathfrak{D}_\phi \subseteqq \mathfrak{D}$ and therefore

$$(4.14) \qquad T_0 \subseteqq H \subseteqq T.$$

Now we shall show that *H is self-adjoint.* For that purpose, we denote for a moment the adjoint operator of H by H^* and the domain of H^* by $\mathfrak{D}_\phi^*$. Then we get from (4. 14)

$$(4.15) \qquad T_0 \subseteqq H^* \subseteqq T.$$

Combined with (4. 2), this shows that $\mathfrak{D}_\phi^*$ consists of all $v \varepsilon \mathfrak{D}$ satisfying

$$(4.16) \qquad [u\bar{v}](b) - [u\bar{v}](a) = 0, \qquad \text{for all } u \varepsilon \mathfrak{D}_\phi.$$

Since the bilinear form $[uv](a)$ is non-singular on $\mathfrak{D}/\mathfrak{D}_a$ and dim $(\mathfrak{D}/\mathfrak{D}_a)$

$= 2\tau_a$, the condition $(4.13)_a$ determines a linear subspace of $\mathfrak{D}$ which is τ_a-dimensional over $\mathfrak{D}_a$, while, since the system of conditions $(4.13)_a$ is assumed to be self-adjoint, τ_a functions ϕ_{aj} $(j = 1, 2, \cdots, \tau_a)$ which are linearly independent $\bmod \mathfrak{D}_a$ satisfy the condition $(4.13)_a$. Hence *the condition* $(4.13)_a$ *is satisfied if and only if u can be represented as*

$$u \equiv \sum_j \gamma^j \phi_{aj} \quad (\bmod \mathfrak{D}_a).$$

Thus we see that *an arbitrary element $u \in \mathfrak{D}$ satisfies the boundary conditions* $(4.13)_a$ *and* $(4.13)_b$ *if and only if u can be represented as*

$$(4.17) \qquad u \equiv \sum_j \gamma^j \phi_{aj} \, (\bmod \mathfrak{D}_a), \qquad \equiv \sum_j \delta^j \phi_{bj} \, (\bmod \mathfrak{D}_b).$$

Moreover, for arbitrary constants $\gamma^1, \gamma^2, \cdots, \delta^1, \delta^2, \cdots$, we can readily construct a function $u \in \mathfrak{D}$ satisfying (4.17). Inserting (4.17) in (4.16) and using (4.11), we infer therefore that v belongs to $\mathfrak{D}_\phi{}^*$ if and only if v satisfies the condition

$$\sum_j \bar{\delta}^j [\phi_{bj} v](b) - \sum_j \bar{\gamma}^j [\phi_{aj} v](a) = 0$$

for arbitrary constants $\gamma^1, \gamma^2, \cdots, \delta^1, \delta^2, \cdots$. Hence $\mathfrak{D}_\phi{}^*$ coincides with $\mathfrak{D}_\phi$. Consequently, comparing (4.15) with (4.14), we see that H^* coincides with H. Thus the operator H is self-adjoint.

DEFINITION 4.2. *The operator H defined above will be called the self-adjoint realization of the formal differential operator L under the system of boundary conditions* $(4.13)_a$ *and* $(4.13)_b$.

5. Spectral theorem. Let H be a fixed self-adjoint realization of the formal differential operator L under the self-adjoint systems of the real boundary conditions

$$(5.1)_a \qquad [\phi_{aj} u](a) = 0, \qquad (j = 1, 2, \cdots, \tau_a),$$

$$(5.1)_b \qquad [\phi_{bj} u](b) = 0, \qquad (j = 1, 2, \cdots, \tau_b),$$

in the sense of Definition 4.2. ϕ_{aj} $(j = 1, \cdots, \tau_a)$ are therefore assumed to be real functions belonging to $\mathfrak{D}$ which are linearly independent $\bmod \mathfrak{D}_a$ and satisfy

$$(5.2)_a \qquad [\phi_{aj} \phi_{ak}](a) = 0, \qquad (j, k = 1, 2, \cdots, \tau_a);$$

similarly ϕ_{bj} $(j = 1, \cdots, \tau_b)$ are real functions belonging to $\mathfrak{D}$ which are linearly independent $\bmod \mathfrak{D}_b$ and satisfy

$$(5.2)_b \qquad [\phi_{bj} \phi_{bk}](b) = 0, \qquad (j, k = 1, 2, \cdots, \tau_b).$$

DEFINITION 5.1. *By the characteristic planes with respect to the operator H, we shall mean the linear spaces $\mathfrak{p}_a(l)$, $\mathfrak{p}_b(l)$ depending on the parameter l defined respectively by*

$$(5.3)_{\mathrm{a}} \qquad \mathfrak{p}_a(l) = \{(f); f \,\varepsilon\, \mathfrak{m}_a(l), [\phi_{aj} w(l,f)](a) = 0 \quad (1 \leqq j \leqq \tau_a)\},$$

$$(5.3)_{\mathrm{b}} \qquad \mathfrak{p}_b(l) = \{(f); f \,\varepsilon\, \mathfrak{m}_b(l), [\phi_{bj} w(l,f)](b) = 0 \quad (1 \leqq j \leqq \tau_b)\},$$

provided $\mathfrak{I}l \neq 0$.

It is to be noted here that the function $w(l,f) = w(x,l,f)$ is not necessarily contained in $\mathfrak{D}$. But, if $f \,\varepsilon\, \mathfrak{m}_a(l)$, for example, then there exists the limit $[\phi_{aj} w(l,f)](a) = \lim [\phi_{aj} w(l,f)](x)$, as one readily infers from Green's formula; thus Definition 5.1 is legitimate. By a similar reason we infer the existence of the limits

$$[w(l,f)\bar{w}(l,g)](a) = \lim_{x \to a} [w(l,f)\bar{w}(l,g)](x), \qquad (f, g \,\varepsilon\, \mathfrak{m}_a(l)),$$

$$[w(l,f)\bar{w}(l,g)](b) = \lim_{x \to b} [w(l,f)\bar{w}(l,g)](x), \qquad (f, g \,\varepsilon\, \mathfrak{m}_b(l)).$$

Incidentally, we conclude, from (2.6), (2.7) and the reality of the boundary conditions $(5.1)_{\mathrm{a}}$, $(5.1)_{\mathrm{b}}$, the relations

$$(5.4) \qquad \mathfrak{p}_a(\bar{l}) = \bar{\mathfrak{p}}_a(l), \qquad \mathfrak{p}_b(\bar{l}) = \bar{\mathfrak{p}}_b(l).$$

THEOREM 5.1. *The characteristic planes $\mathfrak{p}_a(l)$, $\mathfrak{p}_b(l)$ are $(\nu - 1)$-dimensional null planes with respect to l and contained respectively in the sets $\mathfrak{k}_a(l)$, $\mathfrak{k}_b(l)$. Moreover we have*

$$(5.5)_{\mathrm{a}} \qquad [w(l,f)\bar{w}(l,g)](x)/2i\mathfrak{I}l = \int_a^x w(l,f)\bar{w}(l,g)\,dx, \quad (f, g \,\varepsilon\, \mathfrak{p}_a(l)),$$

$$(5.5)_{\mathrm{b}} \quad [w(l,f)\bar{w}(l,g)](x)/2i\mathfrak{I}l = -\int_x^b w(l,f)\bar{w}(l,g)\,dx, \quad (f, g \,\varepsilon\, \mathfrak{p}_b(l)).$$

Proof. We shall prove the theorem with respect to $\mathfrak{p}_a(l)$. For that purpose, we observe the operator L in the interval $(a, c]$, $a < c < b$, instead of (a, b), as in the proof of Theorem 3.2. The meaning of the notations $\mathfrak{H}$, $\mathfrak{D}$, etc. is to be changed correspondingly, so that every $w(x,l,f)$, $f \,\varepsilon\, \mathfrak{m}_a(l)$, becomes a function contained in $\mathfrak{D}$. Now, as was proved in Section 4 above, every function $u \,\varepsilon\, \mathfrak{D}$ satisfying $(5.1)_{\mathrm{a}}$ can be represented as $u \equiv \Sigma_j \gamma^j \phi_{aj}$ (mod $\mathfrak{D}_a$). Hence, if $f \,\varepsilon\, \mathfrak{p}_a(l)$, we have

$$(5.6) \qquad w(l,f) \equiv \sum_j \gamma^j \phi_{aj} \ (\mathrm{mod}\ \mathfrak{D}_a), \qquad (f \,\varepsilon\, \mathfrak{p}_a(l)),$$

while, by hypothesis, ϕ_{aj} $(j = 1, \cdots, \tau_a)$ satisfy $(5.2)_{\mathrm{a}}$. Consequently we get $[w(l,f)w(l,g)] = 0$ for $f, g \,\varepsilon\, \mathfrak{p}_a(l)$, proving that $\mathfrak{p}_a(l)$ is a null plane with respect to l. Again we conclude from (5.6), using the reality of ϕ_{aj} and (4.12), $[w(l,f)\bar{w}(l,g)](a) = 0$ for $f, g \varepsilon \mathfrak{p}_a(l)$. Combined with Green's formula, this yields immediately $(5.5)_a$. Especially for $f = g$, $(5.5)_a$ becomes

$$h(f;x,l) = \int_a^x |w(x,l,f)|^2 dx, \qquad (f \varepsilon \mathfrak{p}_a(l)),$$

proving that $h(f;x,l)$ is positive for every f $(\neq 0)$ lying in $\mathfrak{p}_a(l)$. Hence $\mathfrak{p}_a(l)$ is contained in all $\mathfrak{P}^+(x,l)$, $a < x < b$, and therefore $\mathfrak{p}_a(l) \subset \mathfrak{k}_a(l) = \bigcap_x \mathfrak{P}^+(x,l)$. Finally, since $\mathfrak{p}_a(l)$ is the linear subspace of $\mathfrak{m}_a(l)$ determined by τ_a linear equations $[\phi_{aj}w(l,f)](a) = 0$ $(j = 1, \cdots, \tau_a)$, we have $\dim \mathfrak{p}_a(l) \geqq \dim \mathfrak{m}_a(l) - \tau_a = \nu - 1$. On the other hand, $\mathfrak{k}_a(l)$ has no common point with $\mathfrak{k}_c(l)$ and, as Theorem 2.1 shows, $\mathfrak{k}_c(l)$ contains (at least) one $(\nu - 1)$-dimensional null plane $\mathfrak{p}$ with respect to l. Hence $\mathfrak{p}_a(l)$ has no common point with $\mathfrak{p}$ and therefore

$$\dim \mathfrak{p}_a(l) \leqq \dim \mathfrak{P} - \dim \mathfrak{p} - 1 = \nu - 1.$$

Thus we see that $\dim \mathfrak{p}_a(l)$ is equal to $\nu - 1$.

THEOREM 5.2. *As functions of the parameter l, the characteristic planes $\mathfrak{p}_a(l)$, $\mathfrak{p}_b(l)$ are holomorphic, provided $\Im l \neq 0$, in the sense that, for every l_0 with $\Im l_0 \neq 0$, the suitably normalized*[20] *Plücker's coordinates $\mathfrak{p}_a^{ij\cdots k}(l)$, $\mathfrak{p}_b^{ij\cdots k}(l)$ of $\mathfrak{p}_a(l)$, $\mathfrak{p}_b(l)$ are holomorphic with respect to l in some neighborhoods of l_0.*

Proof. We shall prove the theorem with respect to $\mathfrak{p}_a(l)$. For that purpose, we observe the operator L in the interval $(a, c]$, $a < c < b$, as above. Then we have $\dim \mathfrak{E}(l) = \tau_a + \nu$.

Now let D be an arbitrary compact domain in the half-plane $\Im l > 0$ (or $\Im l < 0$). Then, by virtue of Lemma 3.1, the base $r_\sigma(x,l)$ $(\sigma = 1, 2, \cdots, \tau_a + \nu)$ of the space $\mathfrak{E}(l)$ can be chosen so that $r_\sigma(x,l)$ are holomorphic in D with respect to l in the sense of the norm: $\| \quad \|$ and that $d^m r_\sigma(x,l)/dx^m$ $(m = 0, \cdots, n-1)$ are holomorphic in D with respect to l. Putting $r_\sigma(x,l) = w(x,l,f_\sigma(l))$, we obtain the points $f_\sigma(l)$ $(\sigma = 1, \cdots, \tau_a + \nu)$ generating the space $\mathfrak{m}_a(l)$. Then an arbitrary point f in $\mathfrak{m}_a(l)$ can be represented as $f = \Sigma \eta^\sigma f_\sigma(l)$, so that have $w(x,l,f) = \Sigma \eta^\sigma r_\sigma(x,l)$. Using these η^σ $(\sigma = 1, 2, \cdots, \tau_a + \nu)$ as coordinates of points f in $\mathfrak{m}_a(l)$, we infer from $(5.3)_a$ that the subspace $\mathfrak{p}_a(l)$ of $\mathfrak{m}_a(l)$ is determined by the linear equations

$$(5.7) \qquad \Sigma \eta^\sigma [\phi_{aj} r_\sigma(l)](a) = 0, \qquad (j = 1, 2, \cdots, \tau_a).$$

---

[20] See footnote 14.

These τ_a equations are linearly independent, since $\dim \mathfrak{p}_a(l) = \dim \mathfrak{m}_a(l) - \tau_a$; while the coefficients $[\phi_{aj}r_\sigma(l)](a)$ are holomorphic in D with respect to l, as one readily infers from

$$[\phi_{aj}r_\sigma(l)](a) = [\phi_{aj}r_\sigma(l)](c) - \int_a^c (L[\phi_{aj}] - l\phi_{aj})r_\sigma(l)\,dx.$$

Consequently the ν linearly independent solutions $\eta_\alpha{}^\sigma(l)$ $(\alpha = 1, \cdots, \nu)$ of the equations (5. 7) can be chosen so that $\eta_\alpha{}^\sigma(l)$ are holomorphic in D with respect to l. On the other hand, $f_\sigma(l)$ $(\sigma = 1, 2, \cdots, \tau_a + \nu)$ are also holomorphic in D with respect to l, as was shown in the proof of Theorem 3. 3. Hence putting $g_\alpha(l) = \Sigma_\sigma \eta_\alpha{}^\sigma(l) f_\sigma(l)$, we obtain ν points $g_1(l), \cdots, g_\nu(l)$ depending on l holomorphically in D which generate the characteristic plane $\mathfrak{p}_a(l)$. This proves that $\mathfrak{p}_a(l)$ depends on l holomorphically, provided $\Im l \neq 0$.

After these preparatory considerations, we shall now proceed to the investigation of the spectra of H. First we introduce

DEFINITION 5. 2. *By the characteristic matrix of the operator H with respect to the regular system of fundamental solutions $s_j(x, l)$ $(j = 1, 2, \cdots, n)$ we shall mean the matrix $M^{jk}(l)$ defined by*

$$M^{jk}(l) = M^{jk}(l, \mathfrak{p}_a(l), \mathfrak{p}_b(l)), \qquad (\Im l \neq 0), \tag{5. 8}$$

$\mathfrak{p}_a(l)$, $\mathfrak{p}_b(l)$ being the characteristic planes with respect to H (see (3. 6)).

As (3. 7) shows, *the characteristic matrix $M^{jk}(l)$* is determined by the relation

$$\sum_k \sum_m M^{mj}(l)[s_m(l)s_k(l)] \cdot f^k = \begin{cases} f^j, & (\text{for } f \in \mathfrak{p}_a(l)), \\ 0, & (\text{for } f \in \mathfrak{p}_b(l)). \end{cases} \tag{5. 9}$$

Combined with (5. 4) and (2. 3), this yields

$$M^{jk}(\bar{l}) = \bar{M}^{jk}(l). \tag{5. 10}$$

Again, using (5. 9), we conclude from Theorem 5. 2 that *the characteristic matrix $M^{jk}(l)$ depends on l holomorphically, provided $\Im l \neq 0$.*

DEFINITION 5. 3. *By Green's function for the operator H will be meant the function $G(x, y; l)$ defined as*

$$G(x, y; l) = G(x, y; l, \mathfrak{p}_a(l), \mathfrak{p}_b(l)), \qquad (\Im l \neq 0), \tag{5. 11}$$

$\mathfrak{p}_a(l)$, $\mathfrak{p}_b(l)$ being characteristic planes with respect to H. The integral operator with the kernel $G(x, y; l)$ will be denoted by $G(l)$, i. e., *$G(l) = G(l, \mathfrak{p}_a(l), \mathfrak{p}_b(l))$ (see Definition 3. 2).*

By using the characteristic matrix, Green's function for H can be written in the form

$$(5.12)\qquad G(x,y;l)=G(y,x;l)=\sum_{j,k}M^{jk}(l)s_j(x,l)s_k(y,l),\qquad (x\geqq y),$$

as (3.9) shows. From (5.10) and (2.3) follows therefore

$$(5.13)\qquad G(x,y,\bar{l})=\bar{G}(x,y,l).$$

Now we shall prove

THEOREM 5.3. *The integral operator $G(l)$ coincides with the resolvent $(H-l)^{-1}$ of the operator H, provided $\Im l\neq 0$.*

Proof. Denoting by $\mathfrak{D}_H$ the domain of H, we first prove that, for arbitrary $v\,\varepsilon\,\mathfrak{H}$, $G(l)v$ belongs to $\mathfrak{D}_H$. For that purpose, decompose v into two parts:

$$v=v_0+\bar{w},\qquad (v_0\,\varepsilon\,\mathfrak{H}-\bar{\mathfrak{E}}(l),\ w\,\varepsilon\,\mathfrak{E}(l)).$$

Then, by Theorem 3.5, $G(l)v_0$ belongs to $\mathfrak{D}_0\subset\mathfrak{D}_H$; hence, for the present purpose, it is sufficient to show that $G(l)\bar{w}$ is contained in $\mathfrak{D}_H$. Now we have, by (3.11),

$$G(l)\bar{w}(x)=\sum_{\alpha=1}^{\nu}\sum_{\beta=\nu+1}^{n}F_{\alpha\beta}(l)\{w_\beta(x)\int_a^x w_\alpha\bar{w}\,dx+w_\alpha(x)\int_x^b w_\beta\bar{w}\,dx\},$$

while, using Green's formula, we get

$$2i\Im l\cdot\int_a^x w_\alpha\bar{w}\,dx=[w_\alpha\bar{w}](x)-[w_\alpha\bar{w}](a),$$

$$2i\Im l\cdot\int_x^b w_\beta\bar{w}\,dx=[w_\beta\bar{w}](b)-[w_\beta\bar{w}](x).$$

By virtue of the identity (3.18), we conclude therefore

$$(5.14)\qquad G(l)\bar{w}(x)=(i/2\Im l)\bar{w}(x)$$
$$+(i/2\Im l)\sum_{\alpha=1}^{\nu}\sum_{\beta=\nu+1}^{n}F_{\alpha\beta}(l)\{[w_\alpha\bar{w}](a)\cdot w_\beta(x)-[w_\beta\bar{w}](b)\cdot w_\alpha(x)\}.$$

Let ψ_j $(j=1,\cdots,\tau)$ be real functions belonging to $\mathfrak{D}$ such that

$$\psi_j(x)=\phi_{aj}(x)\ (\text{for } a<x\leqq c),=0\ (\text{for } d\leqq x<b),\qquad (j=1,\cdots,\tau_a),$$

$$\psi_{\tau_a+j}(x)=0\ (\text{for } a\leqq x<c),=\phi_{bj}(x)\ (\text{for } d\leqq x<b),\ (j=1,\cdots,\tau_b),$$

c, d being fixed points such that $a<c<d<b$ (such functions can be readily constructed). Then ψ_j $(j=1,2,\cdots,\tau)$ are linearly independent mod $\mathfrak{D}_0$ and satisfy the conditions

$$(5.15)\qquad [\psi_j\psi_k](b)-[\psi_j\psi_k](a)=0,\qquad (j,k=1,2,\cdots,\tau).$$

By virtue of Theorem 3. 6, ψ_j can be decomposed uniquely as $\psi_j \equiv t_j + \bar{r}_j$ $(\operatorname{mod} \mathfrak{D}_0)$ $(t_j, r_j \varepsilon \mathfrak{E}(l))$, while, since ψ_j are real and $\mathfrak{D}_0 = \bar{\mathfrak{D}}_0$, from the uniqueness of the decomposition it follows that t_j must coincide with r_j. Thus we get

$$(5.16) \qquad \psi_j \equiv r_j + \bar{r}_j \ (\operatorname{mod} \mathfrak{D}_0), \qquad (r_j \varepsilon \mathfrak{E}(l)).$$

Inserting this in (5. 15), we obtain $\Re\{[r_j\bar{r}_k](b) - [r_j\bar{r}_k](a)\} = 0$, while from Green's formula follows $[r_j\bar{r}_k](b) - [r_j\bar{r}_k](a) = 2i\Im l \cdot (r_j, r_k)$. Hence the inner products (r_j, r_k) $(j, k = 1, 2, \cdots, \tau)$ are all real. From this we conclude, using (5. 16), that the functions $r_1, r_2, \cdots, r_\tau$ are linearly independent. Indeed, if $r_1, r_2, \cdots, r_\tau$ were linearly dependent, then there would exist *non-trivial* linear relations $\Sigma_j \eta^j r_j = 0$ with *real* coefficients $\eta^1, \eta^2, \cdots, \eta^\tau$. Whence one would get

$$\Sigma \eta^j \psi_j \equiv \Sigma \eta^j r_j + \Sigma \eta^j \bar{r}_j = 0 \ (\operatorname{mod} \mathfrak{D}_0),$$

contradicting with the linear independence of $\psi_1, \psi_2, \cdots, \psi_\tau \operatorname{mod} \mathfrak{D}_0$. Thus the functions $r_1, r_2, \cdots, r_\tau$ are linearly independent and therefore constitute a base of the space $\mathfrak{E}(l)$. Consequently the function $\bar{w}$ appearing in (5. 14) can be represented as $\bar{w} = \Sigma \gamma^j \bar{r}_j$. Put now $r = \Sigma \gamma^j r_j$. Then we get from (5. 16)

$$(5.17) \qquad r + \bar{w} \equiv \Sigma \gamma^j \psi_j \ (\operatorname{mod} \mathfrak{D}_0), \qquad (r \varepsilon \mathfrak{E}(l)).$$

The functions w_α $(\alpha = 1, \cdots, \nu)$ satisfy the boundary conditions $(5.1)_a$. Hence we have $[\psi_j w_\alpha](a) = 0$ $(j = 1, \cdots, \tau;\ \alpha = 1, 2, \cdots, \nu)$. From (5. 17) follows therefore $[w_\alpha \bar{w}](a) = -[w_\alpha r]$, $(\alpha = 1, 2, \cdots, \nu)$ (remember that $[w_\alpha r](x)$ does not depend on x); similarly we get

$$[w_\beta \bar{w}](b) = -[w_\beta r], \qquad (\beta = \nu + 1, \cdots, n).$$

Inserting these identities in (5. 14) and using (3. 18), we obtain

$$G(l)\bar{w} = (i/2\Im l)\{\bar{w} + r\} \equiv (i/2\Im l)\Sigma \gamma^j \psi_j \ (\operatorname{mod} \mathfrak{D}_0),$$

while it is obvious that ψ_j $(j = 1, \cdots, \tau)$ satisfy the boundary conditions $(5.1)_a$ and $(5.1)_b$. Hence $G(l)\bar{w}$ satisfies also $(5.1)_a$ and $(5.1)_b$, and therefore belongs to $\mathfrak{D}_H$. Thus we conclude that $G(l)v$ belongs to $\mathfrak{D}_H$ for every $v \varepsilon \mathfrak{H}$. This shows, combined with (3. 22), that $(H - l)G(l) = 1$.

To prove the transposed relation

$$(5.18) \qquad G(l)(H - l)u = u, \qquad \text{for all } u \varepsilon \mathfrak{D}_H,$$

we represent, using Theorem 3. 4, every $u \varepsilon \mathfrak{D}_H$ in the form $u = G(l)v + w$, $w \varepsilon \mathfrak{E}(l)$, where $v = (T - l)u$. Then, since $G(l)v$ belongs to $\mathfrak{D}_H$, w is also contained in $\mathfrak{D}_H$ and therefore $Hw = L[w] = lw$ $(\Im l \neq 0)$, whereas the self-adjoint operator H has no complex eigenvalue. Hence w must vanish, so that we obtain $u = G(l)v = G(l)(T - l)u = G(l)(H - l)u$, proving (5. 18), q. e. d.

Now that Theorem 5. 3 has been established, our theory proceeds along the line completely parallel to the case of the differential operator of the second order.[21] The main results can be stated as follows:

THEOREM 5. 4. (SPECTRAL THEOREM). *Let*

$$H = \int_{-\infty}^{+\infty} \lambda dE(\lambda)$$

be the spectral representation of the self-adjoint operator H *and* $M^{jk}(l)$ *the characteristic matrix of* H *with respect to the regular system of fundamental solutions* $s_j(x, l)$ $(j = 1, 2, \cdots, n)$. *Then for every real number* λ, *there exists the limit*

$$\rho^{jk}(\lambda) = \lim_{\delta \to +0} \lim_{\epsilon \to +0} (1/\pi) \int_{\delta}^{\lambda+\delta} \Im M^{jk}(\lambda + i\epsilon)\, d\lambda. \tag{5. 19}$$

The matrix $\rho(\lambda) = (\rho^{jk}(\lambda))$ *thus defined is real and symmetric. As a function of* λ, $\rho(\lambda)$ *is continuous on the right and monotone non-decreasing in the sense that, for* $\mu < \lambda$, *the symmetric matrix* $\rho(\lambda) - \rho(\mu)$ *is positive semi-definite. Put* $E(\Delta) = E(\lambda) - E(\mu)$ *for every finite interval* $\Delta = (\mu, \lambda]$. *Then, for every* $u \in \mathfrak{H}$, $E(\Delta)u(x)$ *can be represented as*

$$E(\Delta)u(x) = \int_a^b u(y)\, dy \int_\Delta \sum_{j,k} s_j(x, \lambda) s_k(y, \lambda)\, d\rho^{jk}(\lambda), \tag{5. 20}$$

where

$$\int_a^b dy \left| \int_\Delta \sum_{j,k} s_j(x, \lambda) s_k(y, \lambda)\, d\rho^{jk}(\lambda) \right|^2 < +\infty,$$

and the integral with respect to y *in* (5. 20) *converges absolutely. Thus* $E(\Delta)$ *is the integral operator with the symmetric kernel*

$$E(x, y; \Delta) = \int_\Delta \sum_{j,k} s_j(x, \lambda) s_k(y, \lambda)\, d\rho^{jk}(\lambda)$$

of Carleman type. $u(x)$ *is represented as*

$$u(x) = \lim_{\lambda \to +\infty,\, \mu \to -\infty} \int_a^b u(y)\, dy \int_\mu^\lambda \sum_{j,k} s_j(x, \lambda) s_k(y, \lambda)\, d\rho^{jk}(\lambda), \tag{5. 21}$$

---

[21] Kodaira, [3].

where the limit converges in the mean. The formula (5.19) *can be rewritten as*

$$(5.22)\qquad \rho^{jk}(\lambda) = \lim_{\delta\to+0}\ \lim_{\epsilon\to+0} (1/\pi)\Im\{\int_{C(\lambda+\delta,\delta,\epsilon)} M^{jk}(l)\,dl\},$$

where $C(\lambda+\delta,\delta,\epsilon)$ *means the oriented polygonal line whose vertices, in order, are* $\delta+i\epsilon$, $\delta+i\alpha$, $\lambda+\delta+i\alpha$, $\lambda+\delta+i\epsilon$, *the real number* α *being subject to the inequality* $\alpha>\epsilon$.[22]

Proof. First we shall treat the special case that the fundamental solutions $s_j(x,l)$ $(j=1,2,\cdots,n)$ constitute the canonical system obtained by solving the differential equation: $L[s]=ls$ under the *real* boundary conditions

$$(5.23)\qquad [d^m s_j(x,l)/dx^m]_{x=c} = e_j^{(m)},\qquad (m=0,\cdots,n-1;\ j=1,\cdots,n)$$

at the fixed point c, $a<c<b$, where $e_j=(e_j^{(0)},e_j^{(1)},\cdots,e_j^{(n-1)})$ are constant real vectors satisfying

$$(5.24)\qquad [e_je_k](c)=\epsilon_{jk}$$

($s_j(x,l)$ mean therefore the solutions $s_j^0(x,l)$ mentioned in Section **2**). We follow the method of H. Weyl.[23] For arbitrary $u.\varepsilon H$, we put

$$u(x,\lambda)=[E(\lambda)-E(0)]u(x),$$

$$u(x,\Delta)=u(x,\lambda)-u(x,\mu)=E(\Delta)u(x),\qquad (\Delta=(\mu,\lambda]);$$

considering $u(x,\Delta)$ as an element of $\mathfrak{H}$, we use the abbreviation $u(\Delta)$ for $u(x,\Delta)$. Obviously $u(\Delta)$ belongs to the domain of H. Putting

$$(5.25)\qquad v(\Delta)=(H-l)u(\Delta)=\int_\Delta(\lambda-l)\,dE(\lambda)u,$$

we have therefore, by Theorem 5.3,

$$(5.26)\qquad u(\Delta)=G(l)v(\Delta).$$

Put, for simiplicity's sake, $G^{(m)}(x,y;l)=\partial^mG(x,y;l)/\partial x^m$, $(y\neq x)$. Then we get from (5.26)

$$(5.27)\qquad \partial^m u(x,\Delta)/\partial x^m=\int_a^b G^{(m)}(x,y;l)\,v(y,\Delta)\,dy,\qquad (m=0,\cdots,n-1),$$

as (3.17) shows. The explicit form of $G^{(m)}(x,y;l)$ is given by

---

[22] In the case of second-order differential operators, this theorem is reduced to the Weyl-Stone-Titchmarsh results. See Weyl, [11], [12], Stone, [7], Theorem 10.22, Titchmarsh, [9], Chap. III. Cf. also Kodaira, [3].

[23] Weyl, [11].

$$(5.28)\qquad G^{(m)}(x,y;l)=\begin{cases}\displaystyle\sum_{\alpha=1}^{\nu}\sum_{\beta=\nu+1}^{n}F_{\alpha\beta}(l)w_\beta{}^{(m)}(x)w_\alpha(y), & (y<x),\\ \displaystyle\sum_{\alpha=1}^{\nu}\sum_{\beta=\nu+1}^{n}F_{\alpha\beta}(l)w_\alpha{}^{(m)}(x)w_\beta(y), & (y>x),\end{cases}$$

whence we see that $G^{(m)}(x,y;l)$ is square summable with respect to y in (a,b), provided $a<x<b$. Thus, *for fixed* x, *the functions* $G^{(m)}(x,y;l)$ *of* y *can be considered as an element of* $\mathfrak{H}$; *then* $G^{(m)}(x,y;l)$ *will be denoted by* $G^{(m)}(x;l)$. By using (5.13), the identity (5.27) can be rewritten as

$$u^{(m)}(x,\Delta)=(v(\Delta),G^{(m)}(x;\bar{l})).$$

Inserting (5.25) for $v(\Delta)$, we obtain from this the important formulae

$$(5.29)\qquad u^{(m)}(x,\Delta)=(u,\int_\Delta(\lambda-\bar{l})dE(\lambda)G^{(m)}(x;\bar{l})),$$

$$(m=0,1,\cdots,n-1).$$

Now, using these formulae, we shall prove that the functions $u^{(m)}(x,\lambda)$ $(m=0,1,\cdots,n-1)$ have the following properties:

(#) *As a function of* λ, *each* $u^{(m)}(x,\lambda)$ *is continuous on the right, of bounded variation in every finite interval* $\Delta=(\mu,\lambda]$ *and, for such intervals* Δ, *the total variation* $\int_\Delta|du^{(m)}(x,\lambda)|$ *is uniformly bounded with respect to* x *in every closed interval* $[x_1,x_2]$, $a<x_1<x_2<b$.

The formulae (5.29) can be rewritten as

$$u^{(m)}(x,\Delta)=(u(\Delta),\int_\Delta(\lambda-\bar{l})dE(\lambda)G^{(m)}(x;\bar{l})),$$

yielding immediately

$$(5.30)\qquad|u^{(m)}(x,\Delta)|\leqq\|u(\Delta)\|\cdot\{\int_\Delta|\lambda-l|^2d\|E(\lambda)G^{(m)}(x;\bar{l})\|^2\}^{1/2}.$$

Consider now decompositions of Δ into the sum of a finite number of subintervals $\Delta_j=(\lambda_{j-1},\lambda_j]$, λ_j being real numbers such that $\lambda_0=\mu<\lambda_1<\cdots<\lambda_h=\lambda$. Then, applying (5.30) to each Δ_j, we get

$$\int_\lambda|du^{(m)}(x,\lambda)|=\sup\sum_j|u^{(m)}(x,\Delta_j)|$$

$$\leqq\|u(\Delta)\|\cdot\{\int_\Delta|\lambda-l|^2d\|E(\lambda)G^{(m)}(x;\bar{l})\|^2\}^{1/2}.$$

Hence, denoting by α the maximum of $|\lambda-l|$ in Δ, we obtain

$$\int_{\Delta} | du^{(m)}(x,\lambda) | \leqq \alpha \, \| u(\Delta) \| \cdot \| G^{(m)}(x;l) \| , \tag{5.31}$$

proving that $u^{(m)}(x,\lambda)$ is a function of bounded variation in Δ, while it is obvious that $u^{(m)}(x,\lambda)$ is continuous on the right (we have naturally presupposed that $E(\lambda)$ is continuous on the right in the sense of the strong topology). Again, as one readily infers by (5.28), $\| G^{(m)}(x;l) \|$ is uniformly bounded with respect to x in every closed interval $[x_1, x_2]$, $a < x_1 < x_2 < b$. Combined with (5.31), this proves that the total variation $\int_{\Delta} | du^{(m)}(x,\lambda) |$ is uniformly bounded in $[x_1, x_2]$. Thus $u^{(m)}(x,\lambda)$ have the properties (#).

As an element of $\mathfrak{H}$, $u(\Delta)$ satisfies obviously the equation

$$Hu(\Delta) = \int_{\Delta} \lambda dE(\lambda) u(\Delta). \tag{5.32}$$

Using this we shall prove that *the function* $u(x,\lambda)$ *satisfies the integro-differential equation*

$$L[u(x,\lambda)] = \int_{+0}^{\lambda} \lambda du(x,\lambda). \tag{5.33}$$

For that purpose, let $\lambda_0, \lambda_1, \cdots, \lambda_h$ be real numbers such that $0 = \lambda_0 < \lambda_1 < \cdots < \lambda_h = \lambda$ and put $\Delta_j = (\lambda_{j-1}, \lambda_j]$, $\delta = \max_j | \lambda_j - \lambda_{j-1} |$. Then we have

$$\Big| \sum_j \lambda_j u(x,\Delta_j) - \int_{+0}^{\lambda} \lambda du(x,\lambda) \Big| \leqq \delta \cdot \int_{+0}^{\lambda} | du(x,\lambda) | .$$

This shows that, in the relation

$$\lim_{\delta \to 0} \sum_j \lambda_j u(x,\Delta_j) = \int_{+0}^{\lambda} \lambda du(x,\lambda), \tag{5.34}$$

the limit converges uniformly with respect to x in every closed interval $[x_1, x_2]$, $a < x_1 < x_2 < b$, since the total variation $\int_{+0}^{\lambda} | du(x,\lambda) |$ is uniformly bounded in $[x_1, x_2]$. On the other hand, applied to $\Delta = (0, \lambda]$, (5.32) means

$$\lim_{\delta \to 0} \int_a^b \Big| L[u(x,\lambda)] - \sum_j \lambda_j u(x,\Delta_j) \Big|^2 dx = 0.$$

Hence we get (5.33).

Incidentally, from (5.34) it follows that $\int_{+0}^{\lambda} \lambda du(x,\lambda)$ is a continuous function of x in $a < x < b$. Combined with (5.33), this shows that, *as a function of* x, $u(x,\lambda)$ *admits in* (a,b) *the continuous* n*-th derivative.*

Now we introduce the functions $u_j(\lambda)$ $(j = 1, 2, \cdots, n)$ of λ associated with $u(x,\lambda)$ by the relations

$$(5.35) \qquad \sum_{j=1}^{n} e_j^{(m)} u^j(\lambda) = u^{(m)}(c, \lambda), \qquad (m = 0, 1, \cdots, n-1),$$

where $e_j^{(m)}$ mean the constants appearing in the boundary conditions (5.23) and therefore $\det(e_j^{(m)}) \neq 0$. Obviously $u^j(\lambda)$ $(j = 1, 2, \cdots, n)$ *are functions of bounded variation in every finite interval* Δ. By using these functions $u_j(\lambda)$, *the function* $u(x, \lambda)$ *can be represented as*

$$(5.36) \qquad u(x, \lambda) = \sum_{j=1}^{n} \int_{+0}^{\lambda} s_j(x, \lambda)\, du^j(\lambda).$$

To prove this we consider the difference

$$t(x, \lambda) = u(x, \lambda) - \sum_{j=1}^{n} \int_{+0}^{\lambda} s_j(x, \lambda)\, du^j(\lambda).$$

As a function of x, $t(x, \lambda)$ admits also continuous derivatives up to the order n and, as a function of λ, $t(x, \lambda)$ is of bounded variation in every finite interval Δ. Furthermore it can be readily verified that $t(x, \lambda)$ satisfies

$$(5.37) \qquad L[t(x, \lambda)] = \int_{+0}^{\lambda} \lambda\, dt(x, \lambda),$$

while, since $s_j^{(m)}(c, \lambda) = e_j^{(m)}$, from (5.35) follows

$$(5.38) \qquad t^{(m)}(c, \lambda) = 0, \qquad (m = 0, 1, \cdots, n-1).$$

Now, by hypothesis, we have

$$\sum_{h+m \leqq n-1} B_{mh}(x) s_j^{(m)}(x, \lambda) s_k^{(h)}(x, \lambda) = [s_j s_k] = \epsilon_{jk},$$

whence we obtain, using (1.6),

$$\sum_{j,k} \epsilon_{jk} s_j^{(m)}(x, \lambda) s_k(x, \lambda) = 0 \ \ (\text{for } 0 \leqq m \leqq n-2), = 1/p_0(x) \quad (\text{for } m = n-1).$$

Hence, putting

$$W(x, y) = \sum_{j,k} \epsilon_{jk} s_j(x, 0) s_k(y, 0),$$

we infer readily that, under the boundary conditions (5.38), the integro-differential equation (5.37) is equivalent to the integral equation

$$t(x, \lambda) = \int_{c}^{x} W(x, y)\, dy \int_{+0}^{\lambda} \lambda\, dw(y, \lambda).$$

Using the method of iteration, we readily infer from this that $t(x, \lambda)$ must vanish identically. Thus (5.36) has been proved.

The explicit forms of $u_j(\lambda)$ are obtained from (5.35) and (5.29). Putting

$$\Gamma^j(x;l) = \begin{cases} \sum\limits_k M^{jk}(l)s_k(x,l), & \text{for } x \leqq c, \\ \sum\limits_k M^{kj}(l)s_k(x,l), & \text{for } x > c, \end{cases} \tag{5.39}$$

we obtain, using (5.12) and (5.23), $G^{(m)}(c,x;l) = \sum\limits_j e_j{}^{(m)}\Gamma^j(x;l)$. From (5.29) follows therefore

$$u^{(m)}(c,\Delta) = \sum_j e_j{}^{(m)}\left(u, \int_\Delta (\lambda - \bar{l})\,dE(\lambda)\Gamma^j(\bar{l})\right). \tag{5.40}$$

Introduce now the functions $\xi^j(\Delta) = \xi^j(x,\Delta)$ defined by

$$\xi^j(\Delta) = \int_\Delta (\lambda - \bar{l})\,dE(\lambda)\Gamma^j(\bar{l}) \tag{5.41}$$

Then, comparing (5.40) with (5.35), we get the important formulae

$$u^j(\Delta) = (u, \xi^j(\Delta)), \qquad (j = 1, 2, \cdots, n). \tag{5.42}$$

As (5.35) shows, $u^j(\Delta)$ can be considered for fixed Δ as a linear functional of $u \in \mathfrak{H}$ *not depending on* l. Hence we infer by (5.42) that *the functions* $\xi^j(\Delta)$ $(j = 1, 2, \cdots, n)$ *are independent of* l, *whereas the definition* (5.41) *of* $\xi^j(\Delta)$ *contains the parameter* l. Obviously $\xi^j(\Delta)$ belongs to $\mathfrak{H}$ and, as (5.41) shows,

$$(\xi^j(\Delta'), \xi^k(\Delta'')) = 0, \qquad \text{if } \Delta' \cap \Delta'' \text{ is empty.} \tag{5.43}$$

The inner products $(\xi^j(\Delta), \xi^k(\Delta))$ *are therefore additive functions of intervals* Δ.

To prove the existence of the limit (5.19), we first define the additive functions $\rho^{jk}(\Delta)$ of intervals Δ by

$$\rho^{jk}(\Delta) = (\xi^j(\Delta), \xi^k(\Delta)), \tag{5.44}$$

and then prove that $\rho^{jk}(\lambda) = \rho^{jk}((0,\lambda])$ are related to $M^{jk}(l)$ by (5.19) (in what follows we assume that the additive functions $\rho^{jk}(\Delta)$, $\rho^{0jk}(\Delta)$ etc. of the intervals Δ are always related to the corresponding functions $\rho^{jk}(\lambda)$, $\rho^{0jk}(\lambda)$, etc. of λ by the equations as follows:

$$\rho^{jk}(\lambda) = \rho^{jk}((0,\lambda]), \qquad \rho^{jk}(\Delta) = \rho^{jk}(\lambda) - \rho^{jk}(\mu) \qquad (\Delta = (\mu,\lambda)).$$

Evidently the matrix $\rho(\Delta) = (\rho^{jk}(\Delta))$ is *hermitian* and *positive semi-definite*; therefore we have

$$\rho^{jj}(\Delta) \geqq 0, \tag{5.45}$$

$$|\rho^{jk}(\Delta)|^2 \leqq \rho^{jj}(\Delta)\rho^{kk}(\Delta). \tag{5.46}$$

From (5.46) follows that $\rho^{jk}(\lambda)$ *are functions of bounded variation in every finite interval* Δ. Now, inserting (5.41) in (5.44), we get

$$\rho^{jk}(\Delta)=\int_{\Delta}|\lambda-l|^{2}(dE(\lambda)\Gamma^{j}(\bar{l}),\Gamma^{k}(\bar{l})),$$

whence we conclude

$$\int_{-\infty}^{+\infty}|\lambda-l|^{-2}d\rho^{jk}(\lambda)=(\Gamma^{j}(\bar{l}),\Gamma^{k}(\bar{l})),\tag{5.47}$$

where the integral converges absolutely, since, by (5.46),

$$\{\int|\lambda-l|^{-2}\,|d\rho^{jk}(\lambda)|\}^{2}\leqq\int|\lambda-l|^{-2}d\rho^{jj}(\lambda)\cdot\int|\lambda-1|^{-2}d\rho^{kk}(\lambda)$$
$$\leqq\|\Gamma^{j}\|^{2}\,\|\Gamma^{k}\|^{2}.$$

The right hand side of (5.47) can be readily calculated. Inserting the explicit form (3.6) of the matrix $M^{jk}(l)$ into (5.39), we get

$$\Gamma^{j}(x;l)=\sum_{\alpha=1}^{\nu}\sum_{\beta=\nu+1}^{n}F_{\alpha\beta}(l)f_{\beta}{}^{j}w(x,l,f_{\alpha}),\qquad(x\leqq c),$$

where f_{α} lie in $\mathfrak{p}_{\alpha}(l)$. Using $(5.5)_{\mathfrak{a}}$, (5.39), (5.23) and (5.24), we obtain therefore

$$2i\Im l\int_{a}^{c}\Gamma^{k}(x;l)\Gamma^{j}(x;\bar{l})\,dx=[\Gamma^{k}(l)\Gamma^{j}(\bar{l})](c-0)$$
$$=\sum_{h}\sum_{m}M^{kh}(l)M^{jm}(\bar{l})\cdot[s_{h}(l)s_{m}(\bar{l})](c)=\sum_{h,m}M^{kh}(l)M^{jm}(\bar{l})\epsilon_{hm};$$

similarly we have

$$2i\Im l\int_{c}^{b}\Gamma^{k}(x;l)\Gamma^{j}(x;\bar{l})\,dx=-\sum_{h,m}M^{hk}(l)M^{mj}(\bar{l})\epsilon_{hm}.$$

Thus we get

$$2i\Im l\cdot(\Gamma^{j}(\bar{l}),\Gamma^{k}(\bar{l}))=\sum_{h,m}\{M^{kh}(l)M^{jm}(\bar{l})-M^{hk}(l)M^{mj}(\bar{l})\}\epsilon_{hm}.$$

Combined with (5.10) and the formula $\Sigma_{h}[M^{hj}(l)-M^{jh}(l)]\epsilon_{hk}=\delta_{k}{}^{j}$, deduced from (3.10), this yields immediately

$$\Im l\cdot(\Gamma^{j}(\bar{l}),\Gamma^{k}(\bar{l}))=\Im M^{jk}(l).$$

Hence we get from (5.47) the *important formulae*

$$\int_{-\infty}^{+\infty}\Im l\cdot|\lambda-l|^{-2}d\rho^{jk}(\lambda)=\Im M^{jk}(l),\tag{5.48}$$

where the integral converges absolutely. Now the relation (5.19) can be readily deduced from (5.48). Indeed we have

$$\lim_{\epsilon\to+0}\int_{\delta}^{\lambda+\delta}\Im M^{jk}(\lambda+i\epsilon)\,d\lambda=\lim_{\epsilon\to+0}\int_{\delta}^{\lambda+\delta}d\lambda\int_{-\infty}^{+\infty}\epsilon[(\mu-\lambda)^2+\epsilon^2]^{-1}d\rho^{jk}(\mu)$$
$$=\int_{-\infty}^{+\infty}d\rho^{jk}(\mu)\lim_{\epsilon\to+0}[\tan^{-1}((\lambda+\delta-\mu)/\epsilon)-\tan^{-1}((\delta-\mu)/\epsilon)]$$
$$=(\pi/2)[\rho^{jk}(\lambda+\delta)+\rho^{jk}(\lambda+\delta-0)-\rho^{jk}(\delta)-\rho^{jk}(\delta-0)],$$

yielding immediately (5.19).

As was already proved, the matrix $\rho(\Delta)=(\rho^{jk}(\Delta))$ is Hermitian, while (5.19) shows that $\rho(\Delta)$ is real; hence $\rho(\Delta)$ must be *symmetric.*

Now, inserting $u=\xi^k(\Delta)$ into (5.42) and (5.36), we get readily the expression

$$(5.49)\qquad \xi^k(x,\Delta)=\int_{\Delta}\sum s_j(x,\lambda)\,d\rho^{kj}(\lambda).$$

Using this, we shall prove the formula (5.20) for $\Delta=(\mu,\lambda]$. For that purpose, let $\lambda_0,\lambda_1,\cdots,\lambda_h$ be real numbers such that $\mu=\lambda_0<\lambda_1<\cdots<\lambda_h=\lambda$, put $\delta=\max_m|\lambda_m-\lambda_{m-1}|$, and consider for fixed x, $a<x<b$, the sum

$$\Xi(y;\delta)=\sum_m\sum_k s_k(x,\lambda_m)\xi^k(y,\Delta_m),\qquad \Delta_m=(\lambda_{m-1},\lambda_m].$$

Choose for every $\epsilon>0$ a positive number $\delta(\epsilon)$ so that $|\lambda'-\lambda''|<\delta(\epsilon)$ implies $|s_k(x,\lambda')-s_k(x,\lambda'')|<\epsilon$ $(k=1,2,\cdots,n)$, provided $\mu\leqq\lambda'<\lambda''\leqq\lambda$. Then, using (5.43), we get the evaluation

$$\int_a^b|\Xi(y;\delta')-\Xi(y;\delta'')|^2dy\leqq 4n\epsilon^2\sum_{k=1}^{n}\|\xi^k(\Delta)\|^2,\qquad(\delta',\delta''<\delta(\epsilon)).$$

This shows that, for $\delta\to0$, the function $\Xi(y;\delta)$ converges to a square summable function $\Xi(y)$ in the mean. The mean convergence implies further the existence of a subsequence $\{\Xi(y,\delta_m)\}$, $\delta_m\to0$, converging to $\Xi(y)$ almost everywhere in (a,b). On the other hand, we infer by (5.49) readily $\lim_{\delta\to0}\Xi(y;\delta)=E(x,y;\Delta)$. Hence $E(x,y;\Delta)$ coincides with $\Xi(y)$ almost everywhere in (a,b); thus we obtain $\int_a^b|E(x,y;\Delta)|^2dy<+\infty$ and

$$(5.50)\qquad \int_a^b|\Xi(y;\delta)-E(x,y;\Delta)|^2dy\to0\qquad(\delta\to0).$$

Now using (5.42) and (5.50), we conclude from (5.36)

$$u(x,\Delta)=\lim_{\delta\to0}\sum_m\sum_j s_j(x,\lambda_m)u^j(\Delta_m)=\lim_{\delta\to+0}(u,\sum_m\sum_j s_j(x,\lambda_m)\xi^j(\Delta_m))$$
$$=\lim_{\delta\to+0}(u,\Xi(\delta))=\int_a^b E(x,y;\Delta)u(y)\,dy,$$

proving (5.20).

Finally the formula (5.21) follows immediately from (5.20), while (5.22) is equivalent to (5.19). Thus the spectral theorem is proved for the special case under consideration.

Now we shall treat the *general case.* For that purpose we denote the special fundamental solutions $s_j(x, l)$ and corresponding $M^{jk}(l), \rho^{jk}(\lambda)$ mentioned above by $s_j^0(x, l)$, $M^{0jk}(l)$, $\rho^{0jk}(\lambda)$, respectively. Then an arbitrary regular system of fundamental solutions $s_j(x, l)$ is obtained from $s_j^0(x, l)$ by a unimodular transformation

$$(U) \qquad s_j(x, l) = \sum_{k=1}^{n} U_j^k(l) s_k^0(x, l),.$$

where $U_j^k(l)$ are entire functions of l satisfying $U_j^k(\bar{l}) = \bar{U}_j^k(l)$. By the transformation (U), the characteristic matrix $M^{jk}(l)$ is transformed as a contravariant tensor, i. e.

$$(5.51) \qquad M^{jk}(l) = \sum_{h,m} V_h^j(l) V_m^k(l) M^{0hm}(l),$$

where $V_j^k(l)$ means the inverse matrix of $U_j^k(l)$; $V_j^k(l)$ are also entire functions of l and satisfy $V_j^k(\bar{l}) = \bar{V}_j^k(l)$.

Let us assume now that $d\rho^{jk}(\lambda)$ constitutes a contravariant tensor, and *define* $\rho^{jk}(\lambda)$ *by the formula*

$$(5.52) \qquad \rho^{jk}(\lambda) = \int_{+0}^{\lambda} \sum_{h,m} V_h^j(\lambda) V_m^k(\lambda) d\rho^{0hm}(\lambda).$$

The matrix $\rho(\lambda) = (\rho^{jk}(\lambda))$ thus defined is also real and symmetric, and, as a function of λ, $\rho(\lambda)$ is continuous on the right and monotone non-decreasing. Furthermore *we have*

$$\int_{\Delta} \sum_{j,k} s_j(x, \lambda) s_k(y, \lambda) d\rho^{jk}(\lambda) = \int_{\Delta} \sum_{j,k} s_j^0(x, \lambda) s_k^0(y, \lambda) d\rho^{0jk}(\lambda),$$

proving the formulae (5.20) and (5.21). To prove the spectral theorem, it is sufficient therefore to verify that the matrix $\rho^{jk}(\lambda)$ thus defined is related to $M^{jk}(l)$ by (5.19).

As to the special matrices $M^{0jk}(l)$, $\rho^{0jk}(\lambda)$, we have the relation (5.48). From this we can deduce the formula

$$(5.53) \qquad M^{0jk}(l) - M^{0jk}(m) = \int_{-\infty}^{+\infty} \{(\lambda - l)^{-1} - (\lambda - m)^{-1}\} d\rho^{0jk}(\lambda) \qquad (\Im l \neq 0, \Im m \neq 0),$$

where the integral converges absolutely.

For an arbitrary positive number σ, we obtain from (5.53) the expression

$$(5.54)\qquad M^{0jk}(l)=\int_{-\sigma}^{\sigma}(\lambda-l)^{-1}d\rho^{0jk}(\lambda)+R_\sigma{}^{0jk}(l),$$

where the *residual term* $R_\sigma{}^{0jk}(l)$ is holomorphic with respect to l except for real l such that $l\leqq-\sigma$ or $l\geqq\sigma$. Corresponding to (5.54), we put

$$(5.55)\qquad M^{jk}(l)=\int_{-\sigma}^{\sigma}(\lambda-l)^{-1}d\rho^{jk}(\lambda)+R_\sigma{}^{jk}(l).$$

Then, inserting (5.54) in (5.51) and using (5.52), we obtain

$$R_\sigma{}^{jk}(l)=\int_{-\sigma}^{\sigma}\sum_{h,m}(\lambda-l)^{-1}\{V_h{}^j(l)V_m{}^k(l)-V_h{}^j(\lambda)V_m{}^k(\lambda)\}d\rho^{0hm}(\lambda)$$
$$+\sum_{h,m}V_h{}^j(l)V_m{}^k(l)R_\sigma{}^{0hm}(l),$$

proving that $R_\sigma{}^{jk}(l)$ is also holomorphic with respect to l except for real l such that $l\leqq-\sigma$ or $l\geqq\sigma$. Furthermore from (5.10) follows

$$(5.56)\qquad R_\sigma{}^{jk}(\bar{l})=\bar{R}_\sigma{}^{jk}(l).$$

To deduce (5.19) from (5.55), choose σ so large that $\sigma>|\lambda|+1$. Then, since $R_\sigma{}^{jk}(l)$ are regular at real μ such that $|\mu|<|\lambda|+1$, we infer, by (5.56), $\Im R_\sigma{}^{jk}(\mu+i\epsilon)\to0$ $(\epsilon\to+0)$ for $|\mu|<|\lambda|+1$. Using this, we conclude

$$\lim_{\epsilon\to+0}\int_\delta^{\lambda+\delta}\Im M^{jk}(\lambda+i\epsilon)d\lambda=\lim_{\epsilon\to+0}\int_\delta^{\lambda+\delta}d\lambda\,\Im\int_{-\sigma}^{\sigma}(\mu-\lambda-i\epsilon)^{-1}d\rho^{jk}(\mu)$$
$$=\tfrac{1}{2}\pi[\rho^{jk}(\lambda+\delta-0)+\rho^{jk}(\lambda+\delta)-\rho^{jk}(\delta-0)-\rho^{jk}(\delta)],$$

yielding immediately (5.19). Thus the spectral theorem is completely proved.

6. Expansion theorem. We introduce for two arbitrary λ-measurable vector functions $\omega(\lambda)=(\omega_1(\lambda),\cdots,\omega_n(\lambda))$, $\chi(\lambda)=(\chi_1(\lambda),\cdots\chi_n(\lambda))$ the notion of their inner product defined by

$$(\omega,\chi)=\int_{-\infty}^{+\infty}\sum_{j,k}\omega_j(\lambda)\bar{\chi}_k(\lambda)d\rho^{jk}(\lambda).$$

Then we have $(\omega,\omega)\geqq0$; $\|\omega\|=(\omega,\omega)^{\frac{1}{2}}$ can be considered therefore as the *norm* of ω. The set Ω of all λ-measurable vector functions ω with $\|\omega\|<+\infty$ constitutes a Hilbert space, if one identifies two functions ω, χ such that $\|\omega-\chi\|=0$. Now we shall prove

THEOREM 6.1. (EXPANSION THEOREM).[24] *For every function $u(x)$ belonging to $\mathfrak{H}$, the integral*

---

[24] Cf. Weyl, [11], [12], Titchmarsh, [9], Chap. III. For series expansions associated with ordinary differential equations of any order, see Birkhoff, [1], Tamarkin, [8].

$$\omega_j(\lambda) = \int_a^b s_j(x, \lambda) u(x)\, dx \tag{6.1}$$

converges in the sense of the norm in Ω *and, by means of* $\omega_j(\lambda)$ *thus defined,* $u(x)$ *is expanded as*

$$u(x) = \int_{-\infty}^{+\infty} \sum_{j,k} s_j(x, \lambda) \omega_k(\lambda)\, d\rho^{jk}(\lambda), \tag{6.2}$$

where the integral converges in the mean; furthermore we have the "Parseval formula"

$$\| \omega \| = \| u \| . \tag{6.3}$$

Conversely, for every $\omega \varepsilon \Omega$, *the integral* (6.2) *converges in the mean and, by means of* $u(x)$ *defined by* (6.2), ω *is represented as* (6.1).

Proof. Choose t, z such that $a < t < z < b$ arbitrarily and put $u^*(x) = u(x)$ (for $t \leqq x \leqq z$), $= 0$ (for $x < t$ or $x > z$). Then the integrals

$$\omega^*_j(\lambda) = \int_t^z s_j(x, \lambda) u(x)\, dx = \int_a^b s_j(x, \lambda) u^*(x)\, dx \tag{6.4}$$

converge absolutely and represent continuous functions of λ. Hence, applying Fubini's theorem, we get, from (5.20),

$$E(\Delta) u^*(x) = \int_\Delta \sum_{j,k} s_j(x, \lambda) \omega^*_k(\lambda)\, d\rho^{jk}(\lambda), \tag{6.5}$$

$$\int_\Delta \omega^*_j(\lambda) \bar{\omega}^*_k(\lambda)\, d\rho^{jk}(\lambda) = \| E(\Delta) u^* \|^2, \tag{6.6}$$

for arbitrary finite intervals $\Delta = (\mu, \lambda]$. Making $\mu \to -\infty$, $\lambda \to +\infty$, we get from (6.6) the identity $\| \omega^* \| = \| u^* \|$; thus the linear transformation $u^* \to \omega^*$ defined by (6.4) is isometric. Now we have $\| u^* - u \| \to 0$ $(t \to a, z \to b)$. Hence there exists an element ω in Ω such that $\| \omega^* - \omega \| \to 0$ $(t \to a, z \to b)$, proving the convergence of the integral (6.1) in the sense of the norm in Ω and the Parseval formula (6.3). To prove that $u(x)$ is represented as (6.2) by means of ω thus defined, we consider the integral

$$I_\Delta(x) = \int_\Delta \sum_{j,k} s_j(x, \lambda) \omega_k(\lambda)\, d\rho^{jk}(\lambda),$$

which evidently converges absolutely. Comparing this with (6.5), we obtain

$$| E(\Delta) u^*(x) - I_\Delta(x) |^2 \leqq \| \omega^* - \omega \|^2 \cdot \int_\Delta s_j(x, \lambda) s_k(x, \lambda)\, d\rho^{jk}(\lambda) ;$$

this proves $| E(\Delta)u^*(x) - I_\Delta(x) | \to 0$ $(t \to a, z \to b)$, where the convergence is uniform in every closed interval $[x_1, x_2]$, $a < x_1 < x_2 < b$. On the other hand, $E(\Delta)u^*$ converges to $E(\Delta)u$ in the mean. Hence we get

$$E(\Delta)u(x) = \int_\Delta \sum_{j,k} s_j(x,\lambda)\omega_k(\lambda)\,d\rho^{jk}(\lambda), \tag{6.9}$$

yielding immediately (6.2).

To prove the converse proposition, put for an arbitrary element ω of Ω

$$\omega_\Delta(\lambda) = \omega(\lambda) \ \ (\text{for } \lambda \varepsilon \Delta), \ = 0 \ \ (\text{for } \lambda \not\varepsilon \Delta),$$

$\Delta = (\mu, \lambda]$ being an arbitrary finite interval. Then the integral

$$u_\Delta(x) = \int_\Delta \sum_{j,k} s_j(x,\lambda)\omega_k(\lambda)\,d\rho^{jk}(\lambda) = \int_{-\infty}^{+\infty} \sum_{j,k} s_j(x,\lambda)\omega_{\Delta k}(\lambda)\,d\rho^{jk}(\lambda) \tag{6.10}$$

converges absolutely and represents a continuous function of x in $a < x < b$. Obviously $\omega_\Delta(\lambda)$ can be approximated by step functions

$$\tilde{\omega}_k(\lambda) = \sum_{m=1}^{h} \gamma_{km} \cdot c(\lambda, \Delta_m)$$

so that $\| \tilde{\omega} - \omega_\Delta \| \to 0$, where Δ_m are mutually disjoint subintervals of Δ and $c(\lambda, \Delta_m)$ denote characteristic functions of Δ_m. As to $\tilde{\omega}$, we have

$$\tilde{u}(x) = \int_{-\infty}^{+\infty} \sum_{j,k} s_j(x,\lambda)\tilde{\omega}_k(\lambda)\,d\rho^{jk}(\lambda) = \sum_m \sum_k \gamma_{km} \cdot \xi^k(x, \Delta_m).$$

Using (5.43) and (5.44), we conclude from this that

$$\| \tilde{u} \|^2 = \sum_m \sum_{j,k} \gamma_{jm}\bar{\gamma}_{km}\rho^{jk}(\Delta_m) = \| \tilde{\omega} \|^2;$$

thus the correspondence $\tilde{\omega} \to \tilde{u}$ is (linear and) isometric. Consequently $\tilde{u}$ converges in the mean to an element u_0 of $\mathfrak{H}$ when $\tilde{\omega}$ approaches to ω_Δ in the sense of the norm. On the other hand, the inequality

$$| \tilde{u}(x) - u_\Delta(x) |^2 \leqq \| \tilde{\omega} - \omega_\Delta \|^2 \cdot \int_\Delta \sum_{j,k} s_j(x,\lambda)s_k(x,\lambda)\,d\rho^{jk}(\lambda)$$

shows that $\tilde{u}(x)$ converges to $u_\Delta(x)$ uniformly in every closed interval $[x_1, x_2]$, $a < x_1 < x_2 < b$, when $\tilde{\omega}$ approaches to ω_Δ. Hence u_Δ coincides with u_0 and therefore $\| u_\Delta \| = \| \omega_\Delta \|$; thus the linear transformation $\omega_\Delta \to u_\Delta$ defined by (6.10) is isometric, proving that the integral (6.2) converges in the mean and that the linear transformation $\omega \to u$ defined by (6.2) is isometric. Now it is obvious that ω is represented as (6.1) by means of $u(x)$ defined by (6.2).

7. Simultaneous differential equations. The above results can be easily extended to the case of simultaneous differential equations. Consider a differential operator

$$L = P_0(x)(d/dx)^h + P_1(x)(d/dx)^{h-1} + \cdots + P_h(x)$$

defined in (a, b) with *matrix* coefficients $P_m(x) = (p_{m\iota\kappa}(x))$, $\iota, \kappa = 1, 2, \cdots, \eta$, operating on vector functions

$$u(x) = (u_1(x), u_2(x), \cdots, u_\eta(x)),$$

where each $p_{m\iota\kappa}(x)$ is a real valued continuous function defined in (a, b) having continuous derivatives up to the order $h - m$ and

$$\det P_0(x) > 0 \ [\text{or} < 0], \qquad (a < x < b). \tag{7.1}$$

The product $n = \eta h$ will be called the "*rank*" of L. Assume furthermore that L coincides with its Lagrange adjoint

$$L^* = (-1)^h (d/dx)^h P_0^*(x) + (-1)^{h-1}(d/dx)^{h-1}P_1^*(x) + \cdots + P_h^*(x),$$

$P_m^*(x)$ being the transposed matrices of $P_m(x)$. In order that L coincides with L^* it is necessary that $P_0(x)$ satisfies $P_0(x) = (-1)^h P_0^*(x)$, which is compatible with (7.1) only if ηh is even. Thus, in the present case, *the order h of L may be odd, whereas the rank $n = \eta h$ must be even:*

$$n = \eta h = 2\nu. \tag{7.2}$$

Now, using the abbreviation

$$u(x)v(x) = \sum_{\kappa=1}^{n} u_\kappa(x) v_\kappa(x),$$

we have *Green's formula*

$$\int_y^x (L[u]v - uL[v])\,dx = [uv](x) - [uv](y), \tag{7.3}$$

where $[uv]$ means the skew-symmetric bilinear form

$$[uv](x) = \sum_{\iota,\kappa} \sum_{j+k \leqq h-1} B_{jk\iota\kappa}(x) u_\iota^{(j)}(x) v_\kappa^{(k)}(x),$$

with the coefficients

$$B_{jk\iota\kappa}(x) = (-1)^{h+j+1} \sum_{m=0}^{h-j-k-1} (-1)^m C_k^{h-j-m-1} p_{m\kappa\iota}^{(h-j-1-k-m)}(x).$$

For $j+k=h-1$, we have

$$B_{jk\iota\kappa}(x) = (-1)^k p_{0\kappa\iota}(x), \qquad (j+k=h-1). \tag{7.4}$$

Hence $[uv](x)$ is a *non-degenerate bilinear form* of two vectors

$$(u_1, u'_1, \cdots, u_1^{(h-1)}, u_2, u'_2, \cdots, u_\eta^{(h-1)}), \qquad (v_1, v'_1, \cdots, v_2, \cdots, v_\eta^{(h-1)}).$$

By means of this bilinear form, we define the "*Wronskian*" of n functions $r, s, t, u, \cdots, v, w$ as

$$[rt\cdots vsu\cdots w] = (1/2^\nu\nu!)\Sigma \pm [rs][tu]\cdots[vw],$$

where $\Sigma\pm$ means the alternating sum extending over all $n!$ permutations of the functions $r, s, t, u, \cdots, v, w$.

Now that it has been established that the bilinear form $[uv](x)$ is non-degenerate, all of our arguments expounded in Sections **2** to **6** can be applied to the present case without essential modifications, since they are entirely based on the bilinear form $[uv](x)$ and Green's formula. It will be sufficient to notice that we have to make the following modifications: First, in Definition 2.1, the functions $s_j(x,l)$ are to be replaced by the vector functions $s_j(x,l) = (s_{j1}(x,l), s_{j2}(x,l), \cdots, s_{j\eta}(x,l))$. Secondly, the inner product (u,v) is to be defined as

$$(u,v) = \int_a^b \sum_\kappa u_\kappa(x)\bar{v}_\kappa(x)\,dx$$

and, correspondingly, $\mathfrak{H}$ will denote the Hilbert space consisting of all measurable *vector functions* u with $(u,u) < +\infty$.

After these modifications, the same arguments as expounded in Sections **2** to **6** lead to the following results: Fix a regular system of fundamental solutions $s_j(x,l)$ $(j=1,2,\cdots,n)$ and associate with every general solution $w(x,l,f) = \Sigma f^j s_j(x,l)$ of the simultaneous differential equations $L[w] = lw$ the point (f) in the $(n-1)$-dimensional projective space $\mathfrak{P}$. Put

$$\mathfrak{m}_b(l) = \{(f)\,;\int_c^b \sum_\kappa |\,w_\kappa(x,l,f)\,|^2 dx < +\infty\},$$

$$\mathfrak{m}_a(l) = \{(f)\,;\int_a^c \sum_\kappa |\,w_\kappa(x,l,f)\,|^2 dx < +\infty\}.$$

Then the dimensions of $\mathfrak{m}_a(l)$ and $\mathfrak{m}_b(l)$ do not depend on l and are not less than $\nu-1$, provided $\Im l \neq 0$. Putting

$$\tau_a = \dim \mathfrak{m}_a(l) - \nu + 1, \qquad \tau_b = \dim \mathfrak{m}_b(l) - \nu + 1,$$

we obtain therefore two non-negative integers τ_a, τ_b which are characteristic for L. Obviously the integers τ_a, τ_b do not exceed ν. Assume now the linearly independent boundary conditions

$$[\phi_{aj}u](a) = 0 \quad (j = 1, 2, \cdots, \tau_a), \qquad [\phi_{bj}u](b) = 0 \quad (j = 1, 2, \cdots, \tau_b),$$

as given, where ϕ_{aj}, ϕ_{bj} are *real* functions belonging to "$\mathfrak{D}$" such that $[\phi_{aj}\phi_{ak}](a) = 0$, $[\phi_{bj}\phi_{bk}](b) = 0$. Then, under these boundary conditions, L becomes a self-adjoint operator, which will be denoted by H. The subspaces

$$\mathfrak{p}_a(l) = \{(f); f \in \mathfrak{m}_a(l), [\phi_{aj}w(l,f)](a) = 0 \ (1 \leqq j \leqq \tau_a)\},$$

$$\mathfrak{p}_b(l) = \{(f); f \in \mathfrak{m}_b(l), [\phi_{bj}w(l,f)](b) = 0 \ (1 \leqq j \leqq \tau_b)\}$$

of $\mathfrak{P}$ will be called the *characteristic spaces* with respect to H. $\mathfrak{p}_a(l)$ and $\mathfrak{p}_b(l)$ have the dimension $\nu - 1$ and have no common point. By the *characteristic matrix* of H we shall mean the matrix $M^{jk}(l)$ defined by the relation

$$\sum_m \sum_j M^{jk}(l)[s_j(l)s_m(l)] \cdot f^m = f^k \text{ (for } f \in \mathfrak{p}_a(l)), = 0 \text{ (for } f \in \mathfrak{p}_b(l)).$$

As a function of l, the matrix $M^{jk}(l)$ is holomorphic in each half plane $\Im l > 0$, $\Im l < 0$, and $M^{jk}(\bar{l}) = \bar{M}^{jk}(l)$. Now we have

THEOREM 7.1. (SPECTRAL THEOREM). *For every real number λ, there exists the limit*

$$\rho^{jk}(\lambda) = \lim_{\delta \to +0} \lim_{\epsilon \to +0} (1/\pi) \int_\delta^{\lambda+\delta} \Im M^{jk}(\lambda + i\epsilon)\, d\lambda.$$

The matrix $\rho(\lambda) = (\rho^{jk}(\lambda))$ thus defined is real and symmetric. As a function of λ, $\rho(\lambda)$ is continuous on the right and monotone non-decreasing in the sense that, for $\mu < \lambda$, the symmetric matrix $\rho(\lambda) - \rho(\mu)$ is positive semi-definite. Let

$$H = \int_{-\infty}^{+\infty} \lambda dE(\lambda)$$

be the spectral decomposition of H and put, for every finite interval $\Delta = (\mu, \lambda]$, $E(\Delta) = E(\lambda) - E(\mu)$. Then $E(\Delta)$ is represented as the integral operator with the matrix kernel

$$E_{\iota\kappa}(x, y, \Delta) = \int_\Delta \sum_{j,k} s_{j\iota}(x, \lambda) s_{k\kappa}(y, \lambda)\, d\rho^{jk}(\lambda)$$

in the sense that

$$E(\Delta)u_\iota(x) = \int_a^b \sum_\kappa E_{\iota\kappa}(x, y, \Delta) u_\kappa(y)\, dy$$

for every $u \in \mathfrak{H}$, *where*

$$\int_a^b \sum_\kappa | E_{\iota\kappa}(x, y, \Delta) |^2 dy < +\infty.$$

Finally, as in **6**, we define the *norm* $\| \omega \|$ of an arbitrary λ-measurable vector function $\omega(\lambda) = (\omega_1(\lambda), \cdots, \omega_n(\lambda))$ by

$$\| \omega \|^2 = \int_{-\infty}^{+\infty} \sum_{j,k} \omega_j(\lambda) \bar{\omega}_k(\lambda) \, d\rho^{jk}(\lambda).$$

Then we have

THEOREM 7. 2. (EXPANSION THEOREM). *For every vector function* $u(x)$ *belonging to* $\mathfrak{H}$, *the integral*

$$\omega_k(\lambda) = \int_a^b \sum_\kappa s_{k\kappa}(x, \lambda) u_\kappa(x) \, dx$$

converges in the sense of the norm defined above and, by means of $\omega_k(\lambda)$ *thus defined, the vector function* $u(x)$ *can be expanded as*

$$u_\kappa(x) = \int_{-\infty}^{+\infty} \sum_{j,k} s_{j\kappa}(x, \lambda) \omega_k(\lambda) \, d\rho^{jk}(\lambda),$$

where the integral converges in the sense of the norm in $\mathfrak{H}$; *furthermore we have the "Parseval formula"*: $\| u \| = \| \omega \|$.

THE INSTITUTE FOR ADVANCED STUDY,
PRINCETON, N. J.

BIBLIOGRAPHY.

1. G. D. Birkhoff, "Boundary value and expansion problems of ordinary linear differential equations," *Transactions of the American Mathematical Society*, vol. 9 (1908), pp. 373-395.
2. I. Halperin, "Closures and adjoints of linear differential operators," *Annals of Mathematics*, vol. 38 (1937), pp. 880-919.
3. K. Kodaira, "Eigenvalue problem for ordinary differential equations of the second order and Heisenberg's theory of S-matrices," *American Journal of Mathematics*, vol. 71 (1949), pp. 921-945.
4. F. J. Murray, *An introduction to linear transformations in Hilbert space*, Princeton (1941).
5. J. von Neumann, "Allgemeine Eigenwerttheorie Hermitescher Funktionaloperatoren," *Mathematische Annalen*, vol. 102 (1930), pp. 49-131.

6. ———, "Ueber adjungierte Funktionaloperatoren," *Annals of Mathematics*, vol. 33 (1932), pp. 294-310.

7. M. H. Stone, *Linear transformations in Hilbert space.* American Mathematical Society Colloquium Publications, vol. 15 (1932).

8. J. Tamarkin, "Some general problems of the theory of ordinary linear differential equations and expansion of an arbitrary function in series of fundamental functions," *Mathematische Zeitschrift*, vol. 27 (1927), pp. 1-54.

9. E. C. Titchmarsh, *Eigenfunction expansions associated with second order differential equations*, Oxford (1946).

10. H. Weyl, "Ueber gewöhnliche lineare Differentialgleichungen mit singulären Stellen und ihre Eigenfunktionen," *Göttinger Nachrichten* (1909), pp. 37-64.

11. ———, "Ueber gewöhnliche Differentialgleichungen mit Singularitäten und die zugehorigen Entwicklungen willkurlicher Funktionen," *Mathematische Annalen*, vol. 68 (1910), pp. 220-269.

12. ———, "Ueber gewöhnliche Differentialgleichungen mit singulären Stellen und ihre Eigenfunktionen," *Göttinger Nachrichten* (1910), pp. 442-467.

13. ———, "Ueber das Pick-Nevanlinna'sche Interpolationsproblem und sein infinitesimales Anagogen," *Annals of Mathematics*, vol. 36 (1935), pp. 230-254.

14. ———, *The classical groups*, Princeton (1939).

ON THE ESSENTIAL SPECTRA OF SINGULAR EIGENVALUE PROBLEMS.*

By PHILIP HARTMAN and AUREL WINTNER.

1. Let λ be a real parameter, and let $p(t) > 0$ and $q(t)$, where $0 \leqq t < \infty$, be a pair of real-valued, continuous functions for which the differential equation

$$(1) \qquad (px')' + (q + \lambda)x = 0,$$

where $' = d/dt$, becomes of *Grenzpunkt* type in the sense of Weyl [4]. This proviso, which will not be repeated below, means that (1) possesses some solution $x(t) = x_\lambda(t)$ violating the L^2-condition

$$(2) \qquad \int_0^\infty x^2(t)\,dt < \infty$$

(for some λ, but then for every λ; cf. [4], p. 238). Correspondingly, (1) and any boundary condition of the form

$$(3_\alpha) \qquad x(0)\cos\alpha + p(0)x'(0)\sin\alpha = 0, \qquad (p(0) > 0),$$

where the angle α is arbitrary, determine an eigenvalue problem.

Let S_α denote the λ-set representing the spectrum of the eigenvalue problem defined by (1) and (3_α). It is known that every S_α is a closed set. According to Weyl ([4], p. 251), the set of the cluster points of S_α is independent of α; it can therefore be denoted simply by S'. The invariant λ-set S' will be referred to as the essential spectrum of (1).

It was proved in [2] that (1) has a non-trivial solution satisfying (2) whenever λ is not in S'. On the other hand, since S' is a closed set, a value $\lambda = \lambda^0$ cannot be in the complement of S' unless the same is true of every λ which is close enough to λ^0. One might expect that these two facts together suffice for the characterization of the complement of S' (and, therefore, of S' itself), but we could not prove this (cf. [3], Appendix). Also that characterization of S' which results, in terms of the zeros of the solutions of

---

* Received November 8, 1949.

(1), from Theorem (I) in [1] is a characterization depending not on a *single* value, λ^0, of λ but is such as to involve an entire neighborhood of λ^0.

In what follows, a theorem will be proved which allows the characterization of a point of S' in such a way as to involve *only that point.* The resulting characterization of points of S' will, in addition, be such as to involve only the differential equation and *no specification of a boundary condition,* (3_α). In fact, the following theorem will be proved:

(*) *A real* $\lambda = \lambda_0$ *is in the essential spectrum of* (1) *if and only if there exists some* $g = g(t)$ *which is real-valued and continuous for* $0 \leqq t < \infty$, *satisfies the* L^2*-condition*

$$\int_0^\infty g^2(t)\,dt < \infty, \tag{4}$$

and is such that the corresponding inhomogeneous differential equation

$$(px')' + (q + \lambda_0)x = g \tag{5}$$

does not possess any solution $x = x(t)$ *satisfying the* L^2*-condition* (2).

2. The proof of this theorem will depend on a Lebesgue-Toeplitz "norm construction," similar to that used in [5] for the characterization of the points of S_α in terms of (5), with a g satisfying (4). The construction applied in [5] will have to be modified so as to replace the spectrum S_α, where α is fixed in (3_α), by the essential spectrum S', which is identical with the λ-set consisting of the common part of every S_α, where $0 \leqq \alpha < \pi$.

The latter characterization of S' is clear from the fact that if $\lambda = \lambda_0$ belongs to two of the spectral sets, say S_α and S_β, where $\alpha \not\equiv \beta \pmod{\pi}$, then $\lambda = \lambda_0$ certainly is a cluster point of at least one of these sets (and, therefore, of both). For otherwise $\lambda = \lambda_0$ is in the point spectrum of two distinct boundary conditions, (3_α) and (3_β). But this is impossible, since (1) does not have (for any λ) two linearly independent solutions, $x = x_\alpha(t)$ and $x = x_\beta(t)$, satisfying (2).

According to Weyl [4], p. 251, the situation belonging to a *fixed* boundary condition is as follows:

(i) A $\lambda = \lambda_0$ is not in S_α if and only if there belongs to every continuous g satisfying (4) *one and only one* solution $x = x_g$ of (5) satisfying (3_α) and (2).

In contrast, (*) can be restated as follows:

(ii) A $\lambda = \lambda_0$ is not in S' if and only if there belongs to every continuous g satisfying (4) *at least one* solution $x = x_g$ of (5) satisfying (2).

Since the complement of the λ-set S' is identical with the sum of the complements of all sets S_α, where $0 \leqq \alpha < \pi$, the assertion of (*) is equivalent to the statement that the elimination of the parameter α from (i) must lead to (ii). In this regard, it is easy to see that the italicized specification in (ii) cannot be replaced by that in (i).

3. The sufficiency of the condition claimed by (*) for a point λ_0 of S' follows at once. For, if there exists a continuous g, satisfying (4), for which (5) fails to have a solution, x, satisfying (2), then it follows from (i) that λ_0 is in S_α, no matter what α may be. In other words, λ_0 is in the common part of all spectra S_α, that is, in S', as claimed by (*).

The converse assertion of (*) is substantially equivalent to the following lemma:

(§) *If* (1) *has no solution* $x(t) \not\equiv 0$ *of class* $L^2(0,\infty)$ *when* $\lambda = \lambda_0$, *then there exists a continuous* $g(t)$, *of class* $L^2(0, \infty)$, *for which* (5) *has no solution* $x(t)$ *of class* $L^2(0,\infty)$.

In fact, if this lemma, (§), is granted, then the necessity of the condition claimed by (*) for a point of S' can be concluded as follows:

Suppose that (5) has, for every continuous $g(t)$ satisfying (4), at least one solution satisfying (2). It then follows from (§) that (1) has, when $\lambda = \lambda_0$, a non-trivial solution of class $L^2(0,\infty)$. Let γ be the angle α (determined mod π) for which this solution satisfies (3_α), and let $x = y(t)$ denote this solution.

For a given g, let $x = x_g(t)$ be a solution of (5) which is of class $L^2(0, \infty)$ (by assumption, there exists such an x_g for every g). Then it is clear from the definition of $y(t)$ that, if c is any constant, the sum $x(t, c) = x_g(t) + cy(t)$ is again an x_g, and all x_g's are of this form. On the other hand, it follows from the definition of the angle γ that, if α is any angle distinct from γ (mod π), then there exists a unique $c = c(\alpha)$ corresponding to which the sum $x(t, c)$ will satisfy (3_α). In fact, since $x(t, c) = x_g(t) + cy(t)$, this $c = c(\alpha)$ is given by the ratio

$$c = -\{x_g(0)\cos\alpha + p(0)x'_g(0)\sin\alpha\}/\{y(0)\cos\alpha + p(0)y'(0)\sin\alpha\}.$$

The denominator is distinct from 0, since, on the one hand, $\alpha \not\equiv \gamma$ (mod π) and, on the other hand, $y(t)$ belongs to the boundary conditions (3_γ).

Accordingly, if $g(t)$ is any continuous function satisfying (4), and if α is any angle distinct from $\gamma \pmod{\pi}$, then (5) has a unique solution satisfying (2) and (3_α). It follows therefore from (i) that λ_0 is not in S_α for any α distinct from $\gamma \pmod{\pi}$. Since S' is the common part of all spectra S_α, this implies that λ_0 cannot be in S', as claimed by (*).

This proves that (*) is true if (§) is granted.

4. The proof of (§) proceeds as follows:

For the given value, λ_0, of λ, choose two (real-valued) linearly independent solutions, say $x = x_1(t)$ and $x = x_2(t)$, of (1). Since their Wronskian, $x_2x_1' - x_1x_2'$, when multiplied by the first coefficient function, $p = p(t)$, of (1), becomes a non-vanishing constant, it can be assumed that

$$p(t)\{x_2(t)x_1'(t) - x_1(t)x_2'(t)\} = 1. \tag{7}$$

On the range $0 \leqq a < b < \infty$, define a function $\mu = \mu(a, b)$ by placing

$$\mu(a, b) = \min_{0 \leqq \theta < \pi} \int_a^b \{x_1(t)\cos\theta + x_2(t)\sin\theta\}^2 d\theta. \tag{8}$$

Then $\mu(a, b)$ is a non-decreasing function of b. Hence, the limit $\mu_a = \mu(a, \infty)$ exists for every non-negative a, with the understanding that the possibility $\mu_a = \infty$ is not excluded. It also follows that $\mu(a, b) \leqq \mu_a$. Actually, $\mu(a, b) < \mu_a$ (whether $\mu_a = \infty$ or $\mu_a < \infty$). In fact, since $x_1(t)$ and $x_2(t)$ are linearly independent, it is clear from (8) that $\mu(a, b)$ is a strictly increasing function of b.

It will be shown that, for every fixed a,

$$\mu(a, b) \to \infty \text{ as } b \to \infty, \tag{9}$$

if λ_0 satisfies the assumption of (§).

Suppose that (9) is false. Then $\mu_a < \infty$ (for some a). But it will be shown that there exists a $\theta^* =$ const. satisfying

$$\int_a^T \{x_1(t)\cos\theta^* + x_2(t)\sin\theta^*\}^2 dt \leqq \mu_a \tag{10}$$

for every $T > a$, and (10) implies that $x(t) = x_1(t)\cos\theta^* + x_2(t)\sin\theta^*$ is of class $L^2(0, \infty)$. In addition, this $x(t)$ is a solution of (1) for $\lambda = \lambda_0$, and it does not vanish identically, since $x_1(t)$ and $x_2(t)$ are linearly independent. Consequently, the assumption of (§) is contradicted.

Accordingly, (9) will be proved if it is shown that there exists a θ^* satisfying (10). But such a θ^* can be constructed as follows:

According to (8), there exists on the interval $0 \leqq \theta < \pi$ at least one $\theta = \theta(a, b)$ for which the integral occurring in (8) attains the value $\mu(a, b)$. Choose $b = n$, where n is any positive integer exceeding a fixed value of a, and put $\theta_n = \theta(a, n)$. Then, if $a < T < n$, it is clear that

$$\int_a^T \{x_1(t)\cos\theta_n + x_2(t)\sin\theta_n\}^2 dt < \mu(a, n) < \mu_a.$$

Hence, in order to conclude the existence of a θ^* satisfying (10), it is sufficient to observe that, since $0 \leqq \theta_n < \pi$, the values θ_n must have at least one cluster value. In fact, the latter can then be chosen to be θ^*.

This concludes the proof of (9).

5. In view of (9) and (8), there exists a sequence of t-values $0 = T_0 < T_1 < \cdots$ satisfying $\mu(T_{n-1}, T_n) > 1$ for $n = 0, 1, \cdots$ and $T_n \to \infty$ as $n \to \infty$. In particular

$$(11) \qquad \int_{T_{n-1}}^{T_n} x_2^2(t)\,dt \geqq \mu(T_{n-1}, T_n) > 1, \text{ where } n = 1, 2, \cdots; T_0 = 0.$$

In fact, the first of the inequalities (11) follows by observing that, in view of the definition (8), the integral of $x_2^2(t)$ over an interval $a \leqq t \leqq b$ cannot be less than $\mu(a, b)$.

In terms of the sequence $T_0, T_1, \cdots$, define for $0 \leqq t < \infty$ a function $g(t)$, which is continuous except for possible jumps at $t = T_n$, by placing

$$(12) \qquad g(t) = C_n x_2(t) \text{ if } T_n \leqq t < T_{n+1},$$

where $n = 0, 1, 2, \cdots$, and, according as $n = 2m$ or $n = 2m + 1$,

$$(13) \qquad 1/C_{2m} = (m + 1)\{\int_{T_{2m}}^{T_{2m+1}} x_2^2(t)\,dt\}^{\frac{1}{2}}, \qquad C_{2m+1} = 0.$$

For this $g(t)$, the value of the integral (2) is seen to be $\Sigma(m + 1)^{-2}$. Hence, this $g(t)$ satisfies (2). It will be shown that, for this $g(t)$, the differential equation (5) fails to have any solution satisfying (2). This will complete the proof of (§), provided that $g(t)$ is allowed to have a sequence of jumps, instead of being continuous as specified in (§). But this complication is immaterial, since it will be clear from the proof below that the above $g(t)$ could readily be smoothed out so as to become continuous.

6. Since $x_1(t)$ and $x_2(t)$ are solutions of the homogeneous case, $g(t) \equiv 0$, of (5), it is readily verified from (7) that, if $g(t)$ is any continuous function,

$$(14) \qquad x(t) = x_1(t) \int_0^t x_2(s) g(s)\, ds - x_2(t) \int_0^t x_1(s) g(s)\, ds$$

is a solution of (5). The same holds if $g(t)$ is not continuous, provided that (5) is replaced (at the discontinuity points, say $t = T_0, T_1, \cdots$, of g) by its integrated form,

$$(5') \qquad p(t)x'(t) - p(0)x'(0) + \int_0^t \{q(s) + \lambda_0\} x(s)\, ds = \int_0^t g(s)\, ds.$$

With this understanding, the general solution of (5) is

$$(15) \qquad x(t) = x_1(t)\{c_1 + \int_0^t x_2(s) g(s)\, ds\} - x_2(t)\{c_2 + \int_0^t x_1(s) g(s)\, ds\},$$

where c_1, c_2 denote arbitrary constants. In fact, the difference of the functions (14), (15) is the general solution of the homogeneous case, $g(t) \equiv 0$, of (5).

Consequently, what remains to be verified can be formulated as follows: If $g(t)$ is defined by (12) and (13), then the function (15) violates the L^2-condition (2) for *every* choice of the constants c_1, c_2.

First, if t is on an interval $T_n \leqq t < T_{n+1}$ belonging to an odd n, then $g(t) = 0$, by (13) and (12). It follows therefore from (15) that

$$(16) \qquad x(t) = A_m x_1(t) - B_m x_2(t) \text{ if } T_{2m+1} \leqq t < T_{2m+2},$$

where A_m, B_m denote the constants

$$(17) \qquad A_m = c_1 + \int_0^{T_{2m+1}} x_2(s) g(s)\, ds, \qquad B_m = c_2 + \int_0^{T_{2m+1}} x_1(s) g(s)\, ds.$$

It is clear from (8) and (16) that

$$(18) \qquad \int_{T_{2m+1}}^{T_{2m+2}} x^2(t)\, dt \geqq (A_m^2 + B_m^2) \mu(T_{2m+1}, T_{2m+2}).$$

On the other hand, the second of the inequalities (11) shows that the product to the right of (18) exceeds $(A_m^2 + B_m^2) \cdot 1$. It follows therefore from (18) that the value of the integral (2) is minorized by the sum of the series

$\Sigma(A_m{}^2+B_m{}^2)$. Consequently, the function (15) cannot satisfy the (L^2)-condition (2) if $\Sigma A_m{}^2=\infty$. Accordingly, the proof will be complete if it is ascertained that $\Sigma A_m{}^2=\infty$ holds for every choice of c_1, c_2 in (15). But this, and even $A_m\rightarrow\infty$ as $m\rightarrow\infty$, can be seen as follows:

According to (12) and (17),

$$A_m=c_1+\sum_{n=1}^{2m+1}C_n\int_{T_{n-1}}^{T_n}x_2{}^2(t)\,dt.$$

It follows therefore from the second of the relations (13) that $A_m\rightarrow\infty$, as $m\rightarrow\infty$, must hold if the series

$$\sum_{m=0}^{\infty}C_{2m}\int_{T_{2m}}^{T_{2m+1}}x_2{}^2(t)\,dt$$

is divergent. But (13) shows that this series must diverge if the series

$$\sum_{m=0}^{\infty}(m+1)^{-1}\{\int_{T_{2m}}^{T_{2m+1}}x_2{}^2(t)\,dt\}^{\frac{1}{2}}$$

does. Since (11) implies that the latter series is minorized by the series $\Sigma(m+1)^{-1}\cdot 1=\infty$, the proof is complete.

7. The content of (*) is that the space of functions g which are real-valued, continuous, satisfy (4) and have the property that (5) possesses no solution satisfying (2), is not empty if and only if $\lambda=\lambda_0$ is in the essential spectrum of (1).

This suggests the consideration of a "dual" space of functions g, namely, the space $\Omega=\Omega(\lambda_0)$ of functions g which are real-valued, continuous, satisfy (4) and have the property that, for *every* α, the differential equation (5) possesses a solution $x=x_{\alpha,g}(t)$ satisfying (2) and (3_α). It turns out that, *for every value of* $\lambda=\lambda_0$, *the linear space* $\Omega(\lambda_0)$ *is not empty and is, in fact,* ∞*-dimensional.*

In order to see this, for a given value of $\lambda=\lambda_0$, let $x_1(t)$ and $x_2(t)$ be two solutions of (1) satisfying (7). If a continuous $g(t)=g_T(t)$, where $T>0$, satisfies $g(t)=0$ for $t\geqq T$ and

$$\int_0^T x_1(t)g(t)\,dt=\int_0^T x_2(t)g(t)\,dt=0, \tag{19}$$

then $g(t)$ belongs to $\Omega(\lambda_0)$. Clearly the set of such functions g is ∞-

dimensional. For a given g, the function $x(t)$ given by (14) is a solution of (5) and satisfies $x(0) = x'(0) = 0$; hence it satisfies (3_α) for every α. On the other hand, if $g(t) = 0$ for $t > T$ and if (19) holds, then $x(t) = 0$ for $t > T$, hence $x(t)$ satisfies (2).

REFERENCES.

[1] P. Hartman, "A characterization of the spectra of one-dimensional wave equations," *American Journal of Mathematics*, vol. 71 (1949), pp. 915-920.

[2] ——— and A. Wintner, "An oscillation theorem for continuous spectra," *Proceedings of the National Academy of Sciences*, vol. 33 (1947), pp. 376-379.

[3] ——— and A. Wintner, "A separation theorem for continuous spectra," *American Journal of Mathematics*, vol. 71 (1949), pp. 650-662.

[4] H. Weyl, "Ueber gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürlicher Funktionen," *Mathematische Annalen*, vol. 68 (1910), pp. 220-269.

[5] A. Wintner, "On the location of continuous spectra," *American Journal of Mathematics*, vol. 70 (1948), pp. 22-30.

THE JOHNS HOPKINS UNIVERSITY.

ON THE EMBEDDING PROBLEM IN DIFFERENTIAL GEOMETRY.*

By Philip Hartman and Aurel Wintner.

---

1. The theorem. Let $g_{ik} = g_{ik}(u, v)$ be the elements of a 2 by 2, symmetric, positive definite (function) matrix, defined in a neighborhood of $(u, v) = (0, 0)$. The problem to be dealt with concerns the existence of a 2-dimensional surface, in a 3-dimensional Euclidean space, for which $ds^2 = g_{11}du^2 + 2g_{12}du\,dv + g_{22}dv^2$; in other words, with the existence of functions $x = x^j(u, v)$, where $j = 1, 2, 3$, which satisfy the three conditions

$$(1) \qquad \sum_{j=1}^{3} (\partial x^j/\partial u^i)(\partial x^j/\partial u^k) = g_{ik}, \qquad (u^1 = u, u^2 = v),$$

in some vicinity of $(0, 0)$. Earlier writers (cf., e. g., [2], pp. 38-39) have assumed that the g_{ik} are analytic functions of (u, v); so that the Cauchy-Kowalewski theorem on partial differential equations becomes applicable. This assumption will not be made below.

If the g_{ik} are of class C^2, then the Theorema Egregium makes it possible to define the curvature $\kappa = \kappa(u, v) = \kappa(u, v; g_{ik})$ belonging to the given g_{ik}. Using the Frobenius representation of the Gaussian curvature, put

$$(2) \qquad \kappa = -\tfrac{1}{4}\Delta^{-4}\, D - \tfrac{1}{2}\Delta^{-1}\{[\Delta^{-1}(g_{11v} - g_{12u})]_v - [\Delta^{-1}(g_{12v} - g_{22u})]_u\},$$

where $\Delta = (g_{11}g_{22} - g_{12}^2)^{\frac{1}{2}} > 0$; the subscripts u, v denote partial differentiation; D is the 3-rowed determinant in which the first column consists of the numbers g_{11}, g_{12}, g_{22}, and the second and third columns are formed by the partial derivatives of g_{11}, g_{12}, g_{22} with respect to u and v, respectively.

A point (u, v) will be called hyperbolic, parabolic or elliptic according as $\kappa(u, v)$ is negative, zero or positive. A (u, v)-domain will be said to consist of "points of the same type" if all of its points belong to one and the same of these three cases.

If $n \geqq 0$ and $0 < \lambda < 1$, a function is called of class $C^n(\lambda)$ in a domain if it possesses n-th order partial derivatives and these partial derivatives satisfy a uniform Hölder condition of order λ with respect to all of the variables.

---

* Received December 6, 1949.

THEOREM. *If $n \geqq 2$ and $0 < \mu < \lambda < 1$, and if the $g_{ik}(u, v)$ are given functions of class $C^n(\lambda)$ and have the property that every point (u, v) in a neighborhood of $(0, 0)$ is of the same type, then there exist three functions $x^j = x^j(u, v)$ of class $C^n(\mu)$ satisfying (1) in a neighborhood of $(0, 0)$.*

The question of existence of functions $x^j = x^j(u, v)$ satisfying (1) is left open in case $(0, 0)$ is a parabolic point which is a cluster point of points of elliptic and/or hyperbolic type.

2. Counter-examples. In a certain sense, the above theorem is the best possible of its type. In fact, if the given functions g_{ik} are of class $C^n(\lambda)$, there need not exist functions $x^j = x^j(u, v)$ of class $C^n(\lambda + \epsilon)$, for any $\epsilon > 0$ satisfying (1). For example, let $n \geqq 2$ and $g_{11} = 1 + v^2 + v^{n+\lambda}$, $g_{12} \equiv 0$, $g_{22} \equiv 1$; so that these functions are of class $C^n(\lambda)$. According to (2), the curvature is

$$\kappa = \tfrac{1}{2} g_{11}^{-1} [2 + c_1 v^{2n+2\lambda-2} + c_2 v^{n+\lambda} + (n + \lambda)(n + \lambda - 1) v^{n+\lambda-2}],$$

where c_1, c_2 are constants. Hence κ is of class $C^{n-2}(\lambda)$, but is not $C^{n-2}(\lambda + \epsilon)$ for any $\epsilon > 0$. On the other hand, if there exist functions $x^j(u, v)$ of class $C^n(\lambda + \epsilon)$, for some $\epsilon > 0$, satisfying (2), it follows from the definition of curvature in terms of the second fundamental form that κ is of class $C^{n-2}(\lambda + \epsilon)$. (Even if $n = 2$, the curvature given by (2) is identical with the Gaussian curvature, defined in terms of the second fundamental form; cf. Section 7 below.)

At first glance, this counter-example seems surprising, since one might expect that, to given g_{ik} of class C^n, there correspond surfaces of class C^{n+1}. The Theorema Egregium indicates, however, the presence of a "paradox," since the curvature, which can be defined in terms of the first and second partial derivatives of the $x^j(u, v)$, is expressed in terms of quantities involving the third order partial derivatives of the $x^j(u, v)$. Actually, a more general formulation of the Theorema Egregium avoids this difficulty. Such a formulation was given by Weyl, [12], pp. 43-44 (cf. van Kampen, [3], p. 135) and will be used in the proof of the Theorem; see Section 7.

3. The parabolic case. If the $g_{ik}(u, v)$ are of class $C^n(\lambda)$ in a neighborhood of $(0, 0)$, then there exists a pair of functions $u = u(u', v')$, $v = v(u', v')$, of class $C^{n+1}(\lambda)$ in a neighborhood of $(u', v') = (0, 0)$, which have the properties that $u(0, 0) = 0$, $v(0, 0) = 0$, that $\partial(u, v)/\partial(u', v') \neq 0$, and that the line-element belonging to

$$(3) \qquad g'_{km}(u', v') = \sum_{j=1}^{2} \sum_{k=1}^{2} g_{ij}(\partial u^i/\partial u'^k)(\partial u^j/\partial u'^m), \qquad (u'^1, u'^2) = (u', v'),$$

is conformal, that is, $g'_{11} = g'_{22}$ and $g'_{12} \equiv 0$. For the case $n = 0$, see Lichtenstein [7]; if $n \geqq 0$, see Lichtenstein [6], also Korn [5]. Clearly, the functions g'_{ik} are of class $C^n(\lambda)$.

If $n \geqq 2$, the invariance of the Gaussian curvature and the Theorema Egregium (that is, (2)) show that

$$\partial^2\Gamma/\partial u'^2 + \partial^2\Gamma/\partial v'^2 = 0, \qquad (e^\Gamma = g'_{11} = g'_{22}),$$

since it is assumed that the curvature is identically 0. Since Γ is of class $C^2(\lambda)$ (hence, of class C^2), it follows that Γ is analytic. Accordingly, there exists, in a neighborhood of $(u^*, v^*) = (0, 0)$, a pair of analytic functions $u = u'(u^*, v^*)$, $v = v'(u^*, v^*)$ having the properties that $u'(0,0) = 0$, $v'(0,0) = 0$, that $\partial(u', v')/\partial(u^*, v^*) \neq 0$, and that the line-element belonging to

$$g^*_{km}(u^*, v^*) = \sum_{j=1}^{2} \sum_{i=1}^{2} g'_{ij}(\partial u'^i/\partial u^{*k})(\partial u'^j/\partial u^{*m}), \qquad (u^{*1}, u^{*2}) = (u^*, v^*).$$

is Euclidean, that is, $g^*_{11} = g^*_{22} \equiv 1$ and $g^*_{12} \equiv 0$.

The composite transformation $(u, v) \rightarrow (u^*, v^*)$ is of class $C^{n+1}(\lambda)$, and has a non-vanishing Jacobian, near $(0, 0)$. Hence the inverse transformation, $u^* = u^*(u, v)$, $v^* = v^*(u, v)$, has the same properties. Put $x^1 = u^*(u, v)$, $x^2 = v^*(u, v)$ and $x^3 \equiv 0$. These functions are of class $C^{n+1}(\lambda)$ and satisfy (1), since $du^{*2} + dv^{*2} = g_{11}du^2 + 2g_{12}du\,dv + g_{22}dv^2$.

This completes the proof of the Theorem in the parabolic case. In fact, the solutions $x^j(u, v)$ of (1) exist, and can be chosen to be of class $C^{n+1}(\lambda)$, in this case (instead of merely of class $C^n(\mu)$, as stated for the general case).

4. Two Lemmas.

Before proceeding to the elliptic and hyperbolic cases, two lemmas concerning the existence of solutions for partial differential equations will be proved. Lemma 2 below is suggested by Picard's theorem [10] on the existence of solutions of linear boundary value problems on small domains and by Lichtenstein's theorem ([8], pp. 89-96) on non-linear elliptic partial differential equations involving a small parameter; principles which go back to H. A. Schwarz. Lemma 2 is also related to the procedure used by Weyl ([12], pp. 64-68) for the "embedding in the large" of a first fundamental form near to that of a sphere.

Let $r > 0$, $0 < \mu < 1$, and let $f(u, v)$ be of class $C^0(\mu)$ (that is, let f satisfy a uniform Hölder condition of order μ) in the circle $\mathcal{B}_r: u^2 + v^2 < r^2$.

By the symbol $|f|_\mu$ will be meant the least (non-negative) number M satisfying both

$$|f(u,v)| \leqq M \text{ and } |f(u+h, v+k) - f(u,v)|(h^2+k^2)^{-\frac{1}{2}\mu} \leqq M$$

for all pairs of distinct points (u,v), $(u+h, v+k)$ in $\mathscr{B}_r$. If a function $w(u,v)$ possesses continuous second order partial derivatives in $\mathscr{B}_r$, put

$$(4) \qquad w_1 = w_{uu}, \quad w_2 = w_{uv}, \quad w_3 = w_{vv}, \quad w_4 = w_u, \quad w_5 = w_v, \quad w_6 = w.$$

If $w(u,v)$ is of class $C^2(\mu)$ in $\mathscr{B}_r$, let $\|w\|$ denote the greatest of the six numbers $|w_1|_\mu, |w_2|_\mu, \cdots, |w_6|_\mu$. For a given f or w of class $C^0(\mu)$ or $C^2(\mu)$ on $\mathscr{B}_R$, the numbers $|f|_\mu$, $\|w\|$ are functions of r (on an interval $0 < r \leqq R$).

LEMMA 1. *In a neighborhood of* $(u,v) = (0,0)$, *let* $a^i = a^i(u,v)$, *where* $i = 1, 2, \cdots, 6$, *be functions of class* C^3 *satisfying*

$$(5) \qquad a^2a^2 - 4a^1a^3 \neq 0,$$

and let $0 < \mu < 1$. *Then there exists a pair of constants* R, M (*depending on* μ) *with the property that if* $0 < r \leqq R$, *and if* $f = f(u,v)$ *is of class* $C^0(\mu)$ *in* $\mathscr{B}_r$, *then there exists a function* $w = w_f = w(u,v)$ *of class* $C^2(\mu)$ *satisfying*

$$(6) \qquad \sum_{i=1}^{6} a^i w_i = f$$

in $\mathscr{B}_r$; *also*

$$(7) \qquad \|w\| \leqq M|f|_\mu;$$

finally, $w = w_f$ *depends linearly on the function* f.

In the applications below, the functions $a^i(u,v)$ will be analytic. In this case, the proof of Lemma 1 shows that if f is of class $C^n(\mu)$, then w_f is of class $C^{n+2}(\mu)$, and that there exists an estimate of the type (7) for higher order partial derivatives.

Proof of Lemma 1 *when* $a^2a^2 - 4a^1a^3 < 0$. In this case, (6) is an inhomogeneous linear partial differential equation of elliptic type. If R is chosen sufficiently small, the function $w = w_f(u,v)$ can be chosen to be the unique solution of class $C^2(\mu)$ of (6) satisfying $w(u,v) = 0$ for $u^2 + v^2 = r^2$; cf. Lichtenstein [8], pp. 91-92 and the references given there. The estimate (7) goes back to Korn [4]. That M can be chosen independent of r (on the range $0 < r \leqq R$) follows from Korn's proof of the existence of M for a fixed r; cf. [7], p. 201, footnote 2.

Proof of Lemma 1 *when* $a^2a^2 - 4a^1a^3 > 0$. In this case, (6) is an inhomogeneous linear differential equation of hyperbolic type. If R is chosen sufficiently small, the function $w = w_f(u, v)$ can be chosen to be the unique solution of class $C^2(\mu)$ of (6) satisfying $w(u, v) = 0$ on suitably fixed arcs through $(0, 0)$. In order to see this, let a transformation $u = u(u', v')$, $v = v(u', v')$, of class C^3 and of non-vanishing Jacobian, be so chosen that, if u, v in (6) are replaced by u', v', then (6), in terms of u', v', appears in the standard normal form

$$w_2 + a^4w_4 + a^5w_5 + a^6w_6 = f_0.$$

If $w = w(u', v')$ denotes that solution of this normal form which satisfies the conditions $w(0, v') \equiv 0$, $w(u', 0) \equiv 0$, then Riemann's integral representation shows that w has the properties claimed in Lemma 1.

LEMMA 2. *In the Monge-Ampère differential equation*

$$a(z_1z_3 - z_2^2) + \sum_{m=1}^{5} \sum_{k=4}^{5} b_{km}z_kz_m + d = 0, \tag{8}$$

let a, b_{km}, d *be functions of* (u, v) *satisfying, in a neighborhood* $(0, 0)$, *a uniform Hölder condition of order* λ, *and let, in that neighborhood,*

$$ad \neq 0. \tag{9}$$

Then, if $0 < \mu < \lambda < 1$, *there exist functions* $z = z(u, v)$ *of class* $C^2(\mu)$ *satisfying* (8) *in some neighborhood of* $(0, 0)$.

The proof will show that if the functions a, b_{km}, d are of class $C^n(\lambda)$ in a neighborhood of $(0, 0)$, then the solution $z = z(u, v)$ can be chosen of class $C^{n+2}(\mu)$.

Remark. That the existence of solutions of (8) in the small is not trivial can be seen from the fact that there exist comparatively innocent-looking partial differential equations possessing *no solution in the small.* For example, if $F(u)$ is a continuous, nowhere differentiable function and $f(u, v) = F(u + v)$, then the considerations of Perron [9], pp. 550-551, show that $z_u - z_v = f$, a linear partial differential equation of first order, has *no* solution.

In order to systematize the notations, the equation (8) will sometimes be written as

$$\sum_{k=1}^{5} \sum_{k=1}^{5} b_{km}z_kz_m + d = 0, \tag{8 bis}$$

where it is supposed that $b_{km} = b_{mk}$ and $b_{11} = b_{12} = b_{23} = b_{33} = 0$, while $2b_{13} = -b_{22} = a$.

Proof of Lemma 2. Denote the expression on the left of (8) by $\phi(u, v, z_1, z_2, \cdots, z_5)$, and consider the partial differential equation

$$\phi(0, 0, z_1, z_2, \cdots, z_5) = 0. \tag{10}$$

Let the numbers $z^0{}_1, z^0{}_3$ be chosen in such a way that $\phi(0, 0, z^0{}_1, 0, z^0{}_3, 0, 0) = 0$; such a choice is clearly possible, since $a(0, 0) \neq 0$ in (8). Since $d(0, 0) \neq 0$, it follows that $z^0{}_1 \neq 0$, $z^3{}_0 \neq 0$. Thus $\partial\phi(0, 0, z^0{}_1, 0, z^3{}_0, 0)/\partial z_1 = a(0, 0)z^0{}_3 \neq 0$. Hence, (10) can be written in the form

$$z_1 - \psi(z_2, z_3, z_4, z_5) = 0 \tag{11}$$

in a neighborhood of $(z_1, \cdots, z_5) = (z^0{}_1, 0, z^0{}_3, 0, 0)$. The function ψ is analytic in its four variables. It follows therefore from the Cauchy-Kowalewski theorem that (11) possesses in a neighborhood of $(0, 0)$ an analytic solution $z = W(u, v)$ satisfying $W_4(0, 0) = W_5(0, 0) = 0$; actually, $W_4(u, 0) = W_u(u, 0)$ and $W_5(u, 0) = W_v(u, 0)$ can be assigned arbitrarily (subject to the restriction that they be analytic, and small in absolute value).

The fact that $z = W$ is a solution of (10) can be written as

$$\sum_{m=1}^{5} \sum_{k=1}^{5} b^0{}_{km} W_k W_m + d^0 = 0, \tag{12}$$

where $b^0{}_{km} = b_{km}(0, 0)$, $d^0 = d(0, 0)$. Put $z = W(u, v) + w$ in (8). This substitution transforms (8) into a partial differential equation for the unknown function w,

$$\sum_{m=1}^{5} \sum_{k=1}^{5} b_{km}(W_k + w_k)(W_m + w_m) + d = 0;$$

cf. (8 bis). Hence, from (12),

$$\sum_{m=1}^{5} \Big(\sum_{k=1}^{5} b^0{}_{km} W_m \Big) w_k = P + \Pi, \tag{13}$$

where

$$P = P(u, v) = \tfrac{1}{2} \sum_{k=1}^{5} \sum_{m=1}^{5} (b^0{}_{km} - b_{km}) W_m W_k + \tfrac{1}{2}(d^0 - d) \tag{14}$$

and

$$\Pi \equiv \Pi(u, v, w_1, \cdots, w_5) = \sum_{m=1}^{5} \sum_{k=1}^{5} (b^0{}_{km} - b_{km}) W_m w_k - \tfrac{1}{2} \sum_{m=1}^{5} \sum_{k=1}^{5} b_{km} w_m w_k. \tag{15}$$

The expression on the left of (13) is of the form (6), where $a^k = \sum_{m=1}^{5} b^0{}_{km} W_m$ for $k = 1, \cdots, 5$, while $a_6 = 0$. Since (8) and (8 bis) are identical, and since $W_4(0, 0) = W_5(0, 0) = 0$, it is seen that

$$a^2a^2 - 4a^1a^3 = 4a^0d^0 \neq 0 \text{ for } (u, v) = (0, 0).$$

Consequently, Lemma 1 is applicable to equations of the form

$$\sum_{k=1}^{5} \left(\sum_{m=1}^{5} b^0{}_{km} W_m \right) w_k = f.$$

The proof of Lemma 2 will now be completed by the use of successive approximations.

5. The Successive Approximations. To this end, a few inequalities will now be collected.

If R is defined as in Lemma 1, let $0 < r < R$. Let the norm symbols $|\cdot\cdot\cdot|_\mu$, $\|\cdot\cdot\cdot\|_\mu$, where $0 < \mu < \lambda < 1$, refer, as above, to the circle $\mathcal{B}_r$: $u^2 + v^2 < r^2$. Let $w(u, v)$, ${}^*w(u, v)$ denote two functions of class $C^2(\mu)$ in $\mathcal{B}_r$ and let α be a number satisfying

$$(16) \qquad \| w \| \leqq \alpha \text{ and } \| {}^*w \| \leqq \alpha.$$

Choose $\beta = \beta(r)$ in such a way that

$$(17) \qquad | b_{km} - b^0{}_{km} |_\mu \leqq \beta \text{ and } | d - d^0 |_\mu \leqq \beta,$$

and that

$$(18) \qquad \beta = \beta(r) \to 0 \text{ as } r \to 0.$$

(that $\beta(r)$ can be chosen to satisfy (18), as well as (17), is a consequence of the assumption that b_{km} and d satisfy a Hölder condition of order $\lambda > \mu$). Finally, let

$$(19) \qquad \Pi(u, v) = \Pi(u, v, w_1, \cdot\cdot\cdot, w_5) \text{ and } {}^*\Pi(u, v) = \Pi(u, v, {}^*w_1, \cdot\cdot\cdot, {}^*w_5).$$

It follows from (14) and (15) that there exist constants A, B such that

$$(20) \qquad | P |_\lambda \leqq A\beta$$

and

$$(21) \qquad | \Pi |_\mu \leqq A\beta(\alpha + \beta) \text{ and } | \Pi - {}^*\Pi |_\mu \leqq B(\alpha + \beta) \| w - {}^*w \|.$$

The constants A, B can be chosen independent of r if $0 < r \leqq r_0 < R$ (for a fixed r_0).

If M denotes the constant occurring in Lemma 1, let $r > 0$ be so small that $1 - MA\beta(r) > 0$, and put

$$\tau = \tau(r) = MA\beta(1 + \beta)/(1 - MA\beta);$$

so that

$$(22) \qquad MA\beta + MA\beta(\tau + \beta) = \tau,$$

and $\tau(r) \to 0$ as $r \to 0$, by (18). If $r > 0$ is sufficiently small, then

$$q = q(r) = MB(\tau + \beta) < 1. \tag{23}$$

For sufficiently small $r > 0$, define a sequence of successive approximations as follows: $w^0(u, v) \equiv 0$ on $\mathscr{B}_r$ and, if $w^0, \cdots, w^j$ have been defined, $w^{j+1}(u, v)$ is that solution of

$$\sum_{k=1}^{5} \Big(\sum_{m=1}^{5} b^0_{km} W_m \Big) w^{j+1}_k = P(u, v) + \Pi(u, v, w^j_1, \cdots, w^j_5) \tag{24}$$

on $\mathscr{B}_r$ which is supplied by Lemma 1.

If $r > 0$ is so small that the above $\tau(r)$ is defined, then

$$\| w^j \| \leqq \tau \text{ for } j = 0, 1, \cdots. \tag{25}$$

In fact, the inequality (25) is trivial for $j = 0$. Suppose that (25) holds for some $j \geqq 0$. Since (20) and (21) are consequences of (16), it follows from (20) and the first inequality in (21) that

$$| P(u, v) + \Pi(u, v, w^j_1, \cdots, w^j_5) |_\mu \leqq A\beta + A\beta(\tau + \beta).$$

Then, from (7) and (22),

$$\| w^{j+1} \| \leqq MA\beta + MA\beta(\tau + \beta) = \tau.$$

This completes the induction proof of (25).

If $j \geqq 1$, equations (24) show that

$$\sum_{k=1}^{5} \Big(\sum_{m=1}^{5} b^0_{km} W_m \Big) (w^{j+1} - w^j)_k$$
$$= \Pi(u, v, w^j_1, \cdots, w^j_5) - \Pi(u, v, w^{j-1}_1, \cdots, w^{j-1}_5).$$

Since the solution w_f, supplied by Lemma 1, depends linearly on f, it follows from (7), and from the second inequality in (21), that

$$\| w^{j+1} - w^j \| \leqq BM(\tau + \beta) \| w^j - w^{j-1} \|.$$

Hence, if $r > 0$ is so small that (23) holds, then the series $\sum \| w^{j+1} - w^j \|$ is convergent. In view of the definition of $\| \cdots \|$, the series $\sum (w^{j+1} - w^j)$ is uniformly convergent, and its sum $w(u, v)$ is of class $C^2(\mu)$ on $\mathscr{B}_r$ and satisfies (13) on $\mathscr{B}_r$.

This completes the proof, since $z = W(u, v) + w(u, v)$ fulfills the assertion of Lemma 2.

Remark. It should be mentioned, for later reference, that $w = w(u, v; r)$ clearly is subject to the inequality $\| w \| \leqq (1 - q)^{-1}\tau$. Hence, $\| w \|$ can be made arbitrarily small by letting $r \to 0$.

6. The elliptic and parabolic cases when $n \geqq 3$. Let the functions $g_{ik} = g_{ik}(u, v)$ be of class $C^n(\lambda)$, where $n \geqq 2$, and let $\kappa(u, v) \neq 0$ in a vicinity $(0, 0)$. Following a formal device of Weingarten ([11], pp. 3-4; cf. Darboux [1], pp. 253-254), the problem will be reduced to the determination of a function $z = z(u, v)$ for which the matrix (γ_{ik}) is positive definite and $\kappa(u, v; \gamma_{ik}) \equiv 0$, where

$$\gamma_{11} = g_{11} - z_u^2, \qquad \gamma_{21} = \gamma_{12} = g_{12} - z_u z_v, \qquad \gamma_{22} = g_{22} - z_v^2.$$

Actually, $\kappa(u, v; \gamma_{ik})$ is not defined unless z is of class C^3. If it is assumed that $z = z(u, v)$ is of class C^3, the equation $\kappa(u, v; \gamma_{ik}) \equiv 0$ becomes a partial differential equation of second order, of the type (8), for z (so that no third order partial derivatives of z appear); cf. Darboux [1], p. 254. By a suitable normalization of (8), the function $a(u, v)$ becomes $-4(g_{11}g_{22} - g_{12}^2) \neq 0$, while $d(u, v)$ is identical with $4(g_{11}g_{22} - g_{12}^2)$. $\kappa(u, v; g_{ik}) \neq 0$. The functions $b_{km}(u, v)$ are quadratic polynomials in the g_{ik} and in their partial derivatives of first and second order.

If the g_{ik} are of class $C^n(\lambda)$, where $n \geqq 2$, then the coefficient functions of (8) are of class $C^{n-2}(\lambda)$. Hence, for sufficiently small r. there exist in $\mathcal{C}_r$ solutions $z = z(u, v)$ of class $C^n(\mu)$. It follows from the Remark at the end of Section **5** that then the matrix (γ_{ik}) is positive definite, if r is sufficiently small. For, on the one hand, $z = W(u, v) + w(u, v)$ and $W_4(0, 0) = W_5(0, 0) = 0$, and, on the other hand, the function $w = w(u, v; r)$ has the property that $\| w \|$ can be made arbitrarily small by choosing r sufficiently small.

Assume that $n \geqq 3$. Then the corresponding functions γ_{ik} are of class $C^{n-1}(\mu)$, and so of class $C^2(\mu)$. An adaptation of the procedure applied in Section **3** shows that there exist functions $u^* = u^*(u, v)$, $v^* = v^*(u, v)$ of class $C^n(\mu)$ satisfying

$$du^{*2} + dv^{*2} = \gamma_{11} + 2\gamma_{12}\, du\, dv + \gamma_{22}\, dv^2.$$

Hence $x^1 = u^*(u, v)$, $x^2 = v^*(u, v)$, $x^3 = z(u, v)$ are of class $C^n(\mu)$ and satisfy (1). This completes the proof of the Theorem (Section **1**) if $n \geqq 3$ and $\kappa \neq 0$.

7. The case $\kappa \neq 0$ and $n = 2$. In order to complete the proof of the theorem for this case, considerations suggested by those of Weyl ([12], pp. 43-44; cf. van Kampen [3], p. 135) will be used.

For given functions g_{ik} of class C^2, the equation (2) and Green's theorem show that

$$(26) \qquad 2\int\!\!\int_T \kappa\Delta\, du\, dv = -\tfrac{1}{2}\int\!\!\int_T \Delta^{-3}D\, du\, dv + \int_S \{\Delta^{-1}(g_{11v} - g_{12u})\,du + \Delta^{-1}(g_{12v} - g_{22u})\,dv\},$$

where S is any simple Jordan curve, of class C^1, and T is its interior. On the other hand, if the g_{ik} are only of class C^1, and if there exists a continuous function $\kappa(u,v)$ satisfying (26), then $\kappa(u,v)$ can be declared to be the curvature belonging to g_{ik}.

If the functions z and g_{ik} are of class C^2, it is easily seen that, in this sense, there exists a curvature $\kappa(u,v;\gamma_{ik})$, and that $\kappa = \phi(u,v,z_1,\cdots,z_5)/4(\det\gamma_{ik})^2$, where ϕ is the expression on the left of (8) as normalized in Section **6**. In fact, (2) shows that it is only necessary to establish the validity of the Green formula

$$-\int_S \delta^{-1}\{(z_vz_{uu} - z_uz_{uv})\,du + (z_vz_{uv} - z_uz_{vv})\,dv$$
$$= 2\int\!\!\int_T \delta^{-1}(z_{uu}z_{vv} - z_{uv})^2 du\, dv$$
$$-\int\!\!\int_T \delta^{-2}\{\delta_v(z_vz_{uu} - z_uz_{uv}) - \delta_u(z_vz_{uv} - z_uz_{vv})\}\,du\, dv,$$

where $\delta = (\det\gamma_{ik})^{\frac{1}{2}}$. Since this formula holds if $z(u,v)$ is of class C^3, it holds also if $z(u,v)$ is of class C^2, as can be seen by approximating z uniformly by polynomials in such a way that the first and second order partial derivatives also converge uniformly.

Let the g_{ik} be given functions of class $C^2(\lambda)$, and let z be of class $C^2(\mu)$ and a solution of (8). Then $\kappa(u,v;\gamma_{ik})$ is defined and is 0. Let $u = u(u',v')$, $v = v(u',v')$ be a transformation of class $C^3(\mu)$ satisfying $u(0,0) = v(0,0) = 0$ and $\partial(u,v)/\partial(u',v') \neq 0$, and having the property that the line-element belonging to

$$\gamma'_{km} = \sum_{j=1}^{2}\sum_{i=1}^{2}\gamma_{ij}(\partial u^i/\partial u'^k)(\partial u^j/\partial u'^m)$$

is conformal, that is, $\gamma'_{11} = \gamma'_{22} > 0$ and $\gamma'_{12} = 0$; cf. Section **3**. Let $\gamma = \gamma'_{11} = \gamma'_{22}$ and $\Gamma = \log\gamma$. The expressions D, Δ in (2), corresponding to conformal γ'_{ik}, are 0, γ, respectively. Since κ is invariant under transformations of class C^3, and since $\kappa \equiv 0$, it follows from (26) that

$$\int_S \{\partial\Gamma/\partial v'\, du' - \partial\Gamma/\partial u'\, dv'\} = 0$$

holds for any sufficiently smooth Jordan curve S. Hence Γ is analytic in a vicinity of $(u', v') = (0, 0)$. The proof can now be completed as in Section **6** (cf. Section **3**).

Remark. An inspection of the partial differential equation (8), occurring in Section **6**, shows that it contains second derivatives of the g_{ik} only in the particular combination which appeared in definition (2) of the curvature κ. Hence the above proof shows that the Theorem (Section **1**) can be refined as follows:

If $n \geqq 2$ *and* $0 < \mu < \lambda < 1$ *and if the* g_{ik} *are of class* $C^n(\lambda)$ *and* κ *is of class* $C^{n-1}(\lambda)$, *then* (1) *possesses solutions* $x^j = x^j(u, v)$ *of class* $C^{n+1}(\mu)$.

For the "parabolic case" of this theorem, cf. the end of Section **3**.

REFERENCES.

[1] G. Darboux, *Leçons sur la théorie générale des surfaces*, vol. 3, Paris, 1894.

[2] M. Janet, "Sur la possibilité de plonger un espace riemannien donné dans un espace euclidien," *Annales de la Société Polonaise de Mathématique*, vol. 5 (1927), pp. 38-43.

[3] E. R. van Kampen, "The theorems of Gauss-Bonnet and Stokes," *American Journal of Mathematics*, vol. 60 (1938), pp. 129-138.

[4] A. Korn, "Ueber Minimalflächen deren Randkurven wenig von ebenen Kurven abweichen," *Abhandlungen der Königlich Preussischen Akademie der Wissenschaften*, Berlin (1909), Anhang II.

[5] ———, "Zwei Anwendungen der Methode der sukzessiven Annäherungen," *Schwarz Festschrift*, Berlin (1914), pp. 215-229.

[6] L. Lichtenstein, "Beweis des Satzes, dass jedes hinreichend kleine, im wesentlichen stetig gekrümmte, singularitätenfreie Flächenstück auf einen Teil einer Ebene zusammenhängend und in den kleinsten Teilen ähnlich abgebildet werden kann," *Abhandlung der Königlich Preussischen Akademie der Wissenschaften*, 1911, Anhang.

[7] ———, "Zur Theorie der konformen Abbildung. Konforme Abbildung nicht-analytischer, singularitätenfreier Flächenstücke auf ebene Gebiete," *Bulletin International de l'Académie des Sciences de Cracovie, Classe des Sciences Mathématiques et Naturelles*, ser. A, 1916, pp. 192-217.

[8] ———, *Vorlesungen über einige Klassen nicht-linearer Integralgleichungen und Integro-Differentialgleichungen*, Berlin (1931).

[9] O. Perron, "Ueber Existenz und Nichtexistenz von Integralen partieller Differentialgleichungssysteme im reellen Gebiet," *Mathematische Zeitschrift*, vol. 27 (1928), pp. 548-564.

[10] E. Picard, "Mémoire sur la théorie des équations aux dérivées partielles et la méthode des approximations successives," *Journal de Mathématiques*, ser. 4, vol. 6 (1890), pp. 145-210.

[11] J. Weingarten, "Ueber die Theorie der auf einander abwickelbaren Oberflächen," *Festschrift der Technischen Hochschule zu Berlin*, 1884.

[12] H. Weyl, "Ueber die Bestimmung einer geschlossenen konvexen Fläche durch ihr Linienelement," *Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich*, vol. 61 (1915), pp. 40-72.

THE JOHNS HOPKINS UNIVERSITY.

A CLOSED SET OF ALGEBRAIC INTEGERS.*

By John B. Kelly.

1. Introduction. A P. V. (Pisot-Vijayaraghavan) number is an algebraic integer, θ, all of whose conjugates over the rational field, with the exception of θ itself, lie inside the unit circle. P. V. numbers are of importance in certain problems of Diophantine approximation, chiefly because 0 and 1 are the only limit points of the sequence whose elements are the fractional parts of the powers of θ. Salem [5], [6], proved the surprising result that the set of P. V. numbers is closed, in the topological sense. Here we shall establish a similar theorem for a closely related class of algebraic integers.

Let us denote by S_1 the class of all P. V. numbers; let us further denote by S_2 the class of all algebraic integers, θ, which have one conjugate, θ' ($\neq \theta$) lying, together with θ, outside the unit circle, while all other conjugates of θ lie inside the unit circle. The numbers of S_2 split naturally into two classes: S'_2, the class of real numbers in S_2 and S''_2, the class of all imaginary numbers in S_2. If θ is in S'_2, θ' is real, while if θ is in S''_2, θ' is imaginary and $\theta' = \bar{\theta}$. S_2, unlike S_1, is not closed, and neither is S'_2. In fact, the set of real quadratic algebraic integers is everywhere dense on the real line. Real quadratic integers outside the unit circle belong either to S_1 or S'_2. The integers of S_1 form a closed set; clearly the quadratic integers of S_1 cannot be everywhere dense. It follows that S'_2 is not closed and that also S_2 is not closed. Now S''_2 is not closed either, for S''_2 may have real limit points; however, the following theorem is true:

Theorem 1. *The imaginary limit points of S''_2 are in S''_2; the real limit points are in S_1. Thus $S_1 + S''_2$ is closed.*

The main part of this paper will be devoted to a proof of Theorem 1.

2. Proof of Theorem 1. Our proof is based upon a succession of lemmas.

In what follows, $c_0, c_1, \cdots, c_n, \cdots$ will be rational integers, while θ will be a complex number such that $|\theta| > 1$. We define the complex numbers $\rho_{1\,n}$ and $\rho_{2\,n}$ by the equations

---

* Received January 12, 1950; revised April 11, 1950.

$$\rho_{1\,n} = \rho_{1\,n}(\theta) = c_n\theta - c_{n+1}, \qquad \rho_{2\,n} = \rho_{2\,n}(\theta) = c_n\theta\bar{\theta} - c_{n+1}(\theta + \bar{\theta}) + c_{n+2}.$$

We have the obvious identity

$$(1) \qquad \rho_{2\,n} = \rho_{1\,n}\bar{\theta} - \rho_{1\,n+1}.$$

LEMMA 1. *If* $\rho_{2\,n}$ *remains bounded as* $n \to \infty$, *then* $c_n = o(|\,\theta\,| + \epsilon)^n$, *where* ϵ *is an arbitrary positive number.*

Proof. Let $|\,\rho_{2\,n}\,| < M$, $n = 0, 1, 2, 3, \cdots$. From (1) we have $|\,\rho_{1\,n+1}\,| < |\,\rho_{1\,n}\,|\,|\,\theta\,| + M$, whence it easily follows, by induction on n, that

$$|\,\rho_{1\,n}\,| \leq |\,\rho_{1\,0}\,|\,|\,\theta\,|^n + M(|\,\theta\,|^n - 1)/|\,\theta\,| - 1 \leq A\,|\,\theta\,|^n,$$

where $A = |\,\rho_{1\,0}\,| + M/|\,\theta\,| - 1$. In other words, $|\,c_n\theta - c_{n+1}\,| \leq A\,|\,\theta\,|^n$, $n = 0, 1, 2, 3, \cdots$. Another induction on n gives $|\,c_n\,| \leq nA\,|\,\theta\,|^{n-1} + |\,c_0\,|\,|\,\theta\,|^n$, from which Lemma 1 follows immediately.

LEMMA 2. *If* $g(z) = \sum_{n=0}^{\infty} a_n z^n$ *is a rational function and* $\lim_{n\to\infty} a_n = 0$, *then* $g(z)$ *has all its poles outside the unit circle.*

Proof. Let $1/\beta_1, 1/\beta_2, \cdots, 1/\beta_k$ be the poles of $g(z)$. Suppose that their orders are $r_1, r_2, \cdots, r_k$, respectively. For all n sufficiently large, $a_n = \sum_{i=1}^{\infty} P_i(n)\beta_i^n$, where $P_i(n)$ is a polynomial of degree $r_i - 1$. Assume that $g(z)$ has poles inside or on the unit circle. Let $1/\beta_1, \cdots 1/\beta_s$, be those of minimum modulus $1/p$, where $p \geq 1$. Since $\lim_{n\to\infty} a_n = 0$, we have $\lim_{n\to\infty} \sum_{i=1}^{s} P_i(n)\beta_i^n = 0$. If $P_i(n) = \sum_{j=0}^{r_i-1} g_{ij}n^j$ and $\beta_i = p\phi_i$ where $|\,\phi_i\,| = 1$, then $\lim_{n\to\infty} p^n \sum_{i=1}^{s} P_i(n)\phi_i^n = \lim_{n\to\infty} p^n \sum_{j=0}^{r} n^j \sum_{i=1}^{s} g_{ij}\phi_i^n = 0$, where $r = \max_i\,(r_i - 1)$. It follows that $\lim_{n\to\infty} \sum_{i=1}^{s} g_{ir}\phi_i^n = \lim_{n\to\infty} \delta_n = 0$. Consider the system of linear equations

$$\sum_{i=1}^{s} g_{ir}\phi_i^{n+p} = \delta_{n+p}, \qquad p = 0, 1, \cdots, s-1.$$

This system has the solution

$$g_{ir} = \sum_{p=0}^{s-1} \lambda_{ip}\delta_{n+p}/\phi_i^n\Delta,$$

where the coefficients λ_{ip} are independent of n while Δ, the Vandermonde determinant derived from $\phi_1, \phi_2, \cdots, \phi_s$, is not zero. Letting $n \to \infty$, we find that $g_{ir} = 0$, $i = 1, 2, \cdots, s$. In the same way $g_{i,r-1} = 0$, $g_{i,r-2} = 0, \cdots$ $g_{i,0} = 0$, $i = 1, 2, \cdots, s$, so that $P_i(n) \equiv 0$, $i = 1, 2, \cdots, s$.

LEMMA 3. *If* $\sum_{n=0} \rho_{2\,n}{}^2$ *converges and if infinitely many of the integers,* c_n, *are distinct from* 0, *then* (a) *if* $\theta \neq \bar{\theta}$, $\theta \in S''_2$, (b) *if* $\theta = \bar{\theta}$, $\theta \in S_1$.

Proof. We show first that $\{c_n\}$ is a recurring sequence, that is, there exist constants $a_1, \cdots, a_k$ such that

$$c_{n+k} + a_1 c_{n+k-1} + \cdots + a_k c_n = 0, \tag{2}$$

for all n greater than some fixed integer n_0. This will follow from the vanishing of the determinant,

$$D_n = \begin{vmatrix} c_0 & c_1 & \cdots & c_n \\ c_1 & c_2 & \cdots & c_{n+1} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ c_n & c_{n+1} & \cdots & c_{2\,n} \end{vmatrix} = \begin{vmatrix} c_0 & c_1 & \rho_{2\,0} & \rho_{2\,1} & \cdots & \rho_{2\,n-2} \\ c_1 & c_2 & \rho_{2\,1} & \rho_{2\,2} & \cdots & \rho_{2\,n-1} \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ c_n & c_{n+1} & \rho_{2\,n} & \rho_{2\,n+1} & \cdots & \rho_{2\,2n-2} \end{vmatrix},$$

for $n > n_0$. (Cf. [1]). From Hadamard's determinant theorem we have

$$D_n{}^2 \leq \left(\sum_{j=0}^{n} c_j{}^2\right)\left(\sum_{j=1}^{n+1} c_j{}^2\right)\left(\sum_{j=0}^{n} \rho_{2j}{}^2\right)\left(\sum_{j=1}^{n+1} \rho_{2j}{}^2\right) \cdots \left(\sum_{j=n-2}^{2n-2} \rho_{2j}{}^2\right).$$

Since, by Lemma 1, $|c_j| < AH^j$, we have $\sum_{j=0}^{n} c_j{}^2 \sum_{j=1}^{n+1} c_j{}^2 < A_1 H^{4n+6}$, where A_1 is a constant. Let $\sum_{j=h}^{\infty} \rho_{2j}{}^2 = R_h$. Then $D_n{}^2 \leq A_1 H^{4n+6} R_0 R_1 R_2 \cdots R_{n-2}$. Since $\lim_{h\to\infty} R_h = 0$, $\lim_{n\to\infty} D_n = 0$. But D_n is an integer. Hence there exists an integer, n_0, such that $D_n = 0$ for all $n > n_0$, and $\{c_n\}$ is a recurring sequence.

Hence $f(z) = \sum_{n=0}^{\infty} c_n z^n$ is a rational function. Also, $g(z) = \sum_{n=0}^{\infty} \rho_{2\,n}(\theta) z^n$ is a rational function and $f(z) = z^2 g(z) + c_0 + (c_1 - (\theta + \bar{\theta})c_0)z/(1 - \theta z)(1 - \bar{\theta} z)$. It follows from Lemma 2 that $g(z)$ has no poles inside the unit circle. $f(z)$ must have poles inside or on the boundary of the unit circle, since otherwise we should have $\lim_{n\to\infty} c_n = 0$. These poles can be only at $1/\theta$ and $1/\bar{\theta}$. By the Fatou-Hurwitz theorem on rational functions whose power series have integral coefficients, cf. [3], the poles of $f(z)$ are the reciprocals of complete sets of conjugate algebraic integers. Thus $\theta \in S''_2$ if $\theta \neq \bar{\theta}$, while $\theta \in S_1$ if $\theta = \bar{\theta}$.

The argument used in obtaining Lemma 3 resembles that employed by Pisot in proving a somewhat similar result for S_1 (Cf. [2], [4]). The obvious analogue of Lemma 3 for S_1 is also true; if $\sum_{n=0}^{\infty} \rho_{1\,n}^2$ converges and if infinitely many of the integers c_n are distinct from 0, then $\theta \varepsilon S_1$.

LEMMA 4. *If $\theta \varepsilon S''_2$ and k is a positive integer, there exist rational integers $c_0, c_1, \cdots, c_n, \cdots$, independent of k, such that*

$$(3) \quad \sum_{n=0}^{k-1} c_n^2 + \sum_{n=k}^{2k-1} (c_n - c_{n-k}(\theta^k + \bar{\theta}^k))^2 + \sum_{n=2k}^{\infty} (c_n - c_{n-k}(\theta^k + \bar{\theta}^k) + c_{n-2k}\theta^k\bar{\theta}^k)^2$$

$$= 1 + (\theta^k + \bar{\theta}^k)^2 + (\theta^k\bar{\theta}^k)^2.$$

Proof. Let θ be a root of the primary irreducible polynomial $P(z)$ which has rational integral coefficients. Let $P(z)$ have degree r and let $Q(z) = z^r P(1/z)$. In some neighborhood of the origin we have $P/Q = \sum_{n=0}^{\infty} c_n z^n$, where the coefficients c_n are rational integers, since $Q(0) = 1$. The function P/Q is regular for $|z| \leqq 1$, save possibly for poles at $1/\theta$ and $1/\bar{\theta}$; and the function $(1 - \theta^k z^k)(1 - \bar{\theta}^k z^k) P/Q$ is regular at all points of $|z| \leq 1$. Setting $z = e^{i\phi}$, we obtain from Parseval's equality

$$\frac{1}{2\pi} \int_0^{2\pi} |(1 - \theta^k z^k)(1 - \bar{\theta}^k z^k) P/Q|^2 \, d\phi = \sum_{n=0}^{k-1} c_n^2 + \sum_{n=k}^{2k-1} (c_n - c_{n-k}(\theta^k + \bar{\theta}^k))^2$$

$$+ \sum_{n=2k}^{\infty} (c_n - c_{n-k}(\theta^k + \bar{\theta}^k) + c_{n-2k}\theta^k\bar{\theta}^k)^2$$

and

$$\frac{1}{2\pi} \int_0^{2\pi} |(1 - \theta^k z^k)(1 - \bar{\theta}^k z^k)|^2 \, d\phi = 1 + (\theta^k + \bar{\theta}^k)^2 + (\theta^k\bar{\theta}^k)^2.$$

Since $|P/Q| = 1$ when $|z| = 1$, these two expressions are equal. Lemma 4 is an extension of a lemma proved by Siegel ([7]) for S_1 and $k = 1$. The same argument is used here. We remark that the function P/Q will actually have poles at $1/\theta$ and $1/\bar{\theta}$ unless $P(1/\theta) = P(1/\bar{\theta}) = 0$. Since θ is in S''_2, this can occur only if θ is a biquadratic unity.

LEMMA 5. *Let $c_0, c_1, \cdots, c_n, \cdots$ be chosen as in Lemma 4. Then, if θ is not a biquadratic unity, c_k and c_{2k} are not both zero, $k = 0, 1, 2, \cdots$.*

Proof. Since $c_0 = P(0)/Q(0) = P(0)$, and $P(z)$ is irreducible, $c_0^2 \geq 1$. Suppose that $c_k = c_{2k} = 0$, for some particular value of $k > 0$. Then the terms on the left-hand side of (3) corresponding to $n = 0$, $n = k$ and $n = 2k$ have the sum $c_0^2(1 + (\theta^k + \bar{\theta}^k)^2 + (\theta^k\bar{\theta}^k)^2)$. This is compatible with Lemma 4 only if $c_0 = \pm 1$, and all remaining coefficients are 0, so that P/Q is regular in $|z| \leq 1$ and θ is a biquadratic unity.

LEMMA 6. *Let* $c_0, c_1, \cdots, c_n, \cdots$, *be chosen as in Lemma* 4. *Then*

$$| c_m | \leq (1 + (\theta^{m+1} + \bar{\theta}^{m+1})^2 + (\theta^{m+1}\bar{\theta}^{m+1})^2)^{1/2} < | \theta |^{2m+2} + 2, \qquad m = 0, 1, 2, \cdots.$$

Proof. Set $k = m + 1$ in Lemma 4 and observe that all terms on the left-hand side of (3) are non-negative.

We turn now to the proof of Theorem 1. Let α be the limit of a sequence $\{\theta_s\}$ of algebraic integers in S''_2. We are to show that α is either in S''_2 or in S_1.

Proof. The sequence $\{\theta_s\}$ is bounded; for all s we have $| \theta_s | < M$. Hence the sequence $\{\theta_s\}$ can contain only a finite number of biquadratic unities, for the number of algebraic integers of a given degree, lying, together with their conjugates in a bounded region, is finite. We may therefore suppose that these biquadratic unities are deleted from $\{\theta_s\}$.

Corresponding to each element, θ_s, of our sequence, we construct, as in Lemma 4, a sequence of integers c_{ns} so that

$$(4) \quad c_{0\,s}^2 + (c_{0\,s}(\theta_s + \bar{\theta}_s) - c_{1\,s})^2 + \sum_{n=0}^{\infty} \rho_{2\,n}^2(\theta_s) = 1 + (\theta_s + \bar{\theta}_s)^2 + (\theta_s\bar{\theta}_s)^2.$$

From Lemma 6 we have

$$(5) \qquad | c_{n\,s} | < M^{2n+2} + 2, \qquad n = 0, 1, 2, \cdots, s = 1, 2, 3, \cdots.$$

The bounded sequence $c_{0\,s}$ consists only of integers; hence some integers must occur infinitely often. Call one of these integers c_0. We may extract an infinite subsequence $\{\theta_{s_0}\}$ from our original sequence such that $c_{0\,s_0} = c_0$, $s_0 = 1, 2, 3, \cdots$. Clearly $c_0 \neq 0$.

Similarly, the bounded sequences $c_{1\,s_0}$ and $c_{2\,s_0}$, consist only of integers, so that some integers must occur infinitely often in each of these sequences. Choose one of these integers from each sequence and call them c_1 and c_2. We infer from Lemma 5 that one of the sequences $c_{1\,s_0}$ and $c_{2\,s_0}$ contains infinitely many elements distinct from 0, hence at least one of the integers c_1, c_2, may be chosen distinct from 0. We may extract an infinite subsequence $\{\theta_{s_1}\}$ from the sequence $\{\theta_{s_0}\}$ such that $c_{0\,s_1} = c_0$, $c_{1\,s_1} = c_1$, $c_{2\,s_1} = c_2$, $s_1 = 1, 2, 3, \cdots$.

Proceeding in this way, we construct an infinite sequence of integers: $\{c_n\}$. Repeated application of Lemma 5 enables us to choose the integers c_q and c_{2q} (where $2q$ contains 2 to an odd power) in such a way that one of them is distinct from 0. Moreover, for any integer, $m \geq 0$, there exists an infinite subsequence $\{\theta_{s_m}\}$ of our original sequence $\{\theta_s\}$ such that $c_{n\,s_m} = c_n$, $n = 0, 1, 2, \cdots, m$, $s_m = 1, 2, 3, \cdots$.

We now demonstrate the convergence of the series $\sum_{n=0}^{\infty} \rho_{2n}^2(\alpha)$. Theorem 1 will then follow from Lemma 3, provided that $|\alpha| \neq 1$. We have

$$(6) \qquad \sum_{n=0}^{m-2} \rho_{2n}^2(\alpha) = \sum_{n=0}^{m-2} [\rho_{2n}^2(\alpha) - \rho_{2n}^2(\theta_{s_m})] + \sum_{n=0}^{m-2} \rho_{2n}^2(\theta_{s_m}).$$

Let ϵ be an arbitrary positive number. Choose s_m so large that $|\alpha - \theta_{s_m}| < \epsilon$. Observing that $c_n = c_{n\,s_m}$, $n = 0, 1, 2, \cdots, m$, we find that

$$|\rho_{2n}(\alpha) - \rho_{2n}(\theta_{s_m})| \leq |c_n| [|\alpha|^2 - |\theta_{s_m}|^2] + 2 |c_{n+1}| |\alpha - \theta_{s_m}|,$$
$$n = 0, 1, \cdots, m - 2.$$

From (5) we obtain $|c_n| = |c_{n\,s_m}| < M^{2m+2} + 2$, $n = 0, 1, 2, \cdots, m$, whence

$$(7) \qquad |\rho_{2n}(\alpha) - \rho_{2n}(\theta_{s_m})| < (M^{2m+2} + 2)(2M + 2)\epsilon,$$
$$n = 0, 1, 2, \cdots, m - 2.$$

Similarly

$$(8) \qquad |\rho_{2n}(\alpha) + \rho_{2n}(\theta_{s_m})| < 2(M^{2m+2} + 2)(M + 1)^2,$$
$$n = 0, 1, 2, \cdots, m - 2.$$

Multiplying (8) and (7) and summing over n, we have

$$\left|\sum_{n=0}^{m-2} [\rho_{2n}^2(\alpha) - \rho_{2n}^2(\theta_{s_m})]\right| < 4(M^{2m+2} + 2)^2 (M + 1)^3 (m - 1)\epsilon,$$
$$n = 0, 1, 2, \cdots, m - 2.$$

Since ϵ is arbitrary, we can conclude from (6) and (4) that $\sum_{n=0}^{m-2} \rho_{2n}^2(\alpha) \leq 1 + 4M^2 + M^4$. Consequently, $\sum_{n=0}^{\infty} \rho_{2n}^2(\alpha)$ is convergent.

We complete the proof by showing that the excluded case $|\alpha| = 1$ cannot arise.

Any positive integral power of a number in S''_2 is either in S''_2 or in S_1. If the sequence $\{\theta_s\}$ converges to $\alpha = e^{i\phi}$, then the set of all positive integral powers of all numbers in the sequence will have a limit point, α_R, on each circle $r = R$, $R \geq 1$. If $R > 1$, then α_R must, by what we have already proved, belong to S''_2 or S_1. Every number in S''_2 or S_1 is algebraic and must have an algebraic modulus. But R may be transcendental. This contradiction shows that $|\alpha| \neq 1$.

3. Concluding remarks. We could have dispensed with Lemma 1 altogether in our proof of Theorem 1 if we had strengthened the hypothesis in Lemma 3 by requiring that $c_n = O(B^n)$, for some $B > 1$. This condition is fulfilled by the integers c_n used in the proof of Theorem 1, on account of Lemma 6. However, we feel that Lemma 3 is of some interest in its own right, and we have stated it in as strong a form as possible.

Either of Salem's proofs that S_1 is closed may be employed, with more or less obvious modifications, to prove that part of Theorem 1 which concerns the imaginary limit points of S''_2; however, we were unsuccessful in our attempts to adapt them to the rest of Theorem 1. The basic idea underlying our proof is similar to that underlying his first proof [5], but there are considerable differences in detail. Our Lemma 3 plays the rôle of Pisot's theorem, while Lemma 4 play the rôle of Salem's lemma.

We have tried to extend Theorem 1 to still wider classes of algebraic integers. Let S_k be the class of all algebraic integers, θ, having exactly k conjugates, including θ itself, outside the unit circle, and all remaining conjugates inside the unit circle, with none *on* the circle. An argument just like that given in the introduction for $k = 2$ shows that S_k is not closed if $k > 1$. If, ignorant of this fact, we were to try to use the methods of this paper to prove that S_k is closed, we should find that, although our sequence $\{\theta_s\}$ is bounded, the conjugates $\theta_s^{(i)}$, $i = 1, 2, \cdots, k-1$, $s = 1, 2, \cdots$ may very well be unbounded. It is natural then, to impose the further restriction that the conjugates be bounded also. The right-hand side of the analogue of Lemma 4 with θ replaced by θ_s, is then a bounded function of s, a property which we used several times in our proof of Theorem 1. We can then extract a subsequence of the sequence $\{\theta_s\}$ such that we can form $k-1$ sequences of conjugates which are also convergent with limits, say, $\alpha_1, \alpha_2, \cdots, \alpha_{k-1}$. One might expect that then each of the numbers $\alpha, \alpha_1, \alpha_2, \cdots, \alpha_{k-1}$ would belong to some S_r with $r \leq k$. However, one finds that the complete analogue of Lemma 3 is false. All that we can say at the present time is that at least one of the numbers $\alpha, \alpha_1, \cdots, \alpha_{k-1}$ belongs to some S_r, with $r \leq k$.

Since $S_1 + S''_2$ is closed, every point of $S_1 + S''_2$ is either a limit point or an isolated point. It would be of interest to determine which points of $S_1 + S''_2$ are limit points and which are isolated points. Siegel ([7]) has carried through a similar investigation for S_1; he has shown that S_1 has no limit points in the closed interval, $[1, 2^{\frac{1}{2}}]$ and has found all the isolated points in this interval. However, nothing is known beyond this; the identity of the smallest limit point has not yet been determined. The problem for $S_1 + S''_2$ appears to be even more difficult. We can merely state that, as shown at the conclusion of the proof of Theorem 1, $S_1 + S''_2$ has no limit points on the unit circle.

The author desires to thank the referee for helpful criticisms which led to certain simplifications in the presentation of the argument.

UNIVERSITY OF WISCONSIN.

REFERENCES.

[1] L. Bieberbach, *Lehrbuch der Funktionentheorie*, vol. 2 (1931), pp. 319-321.

[2] C. Pisot, "La répartition mod. 1 et les nombres algébriques," *Annali di Pisa* (2), vol. 7 (1938), pp. 205-248.

[3] G. Pólya-G. Szegö, *Aufgaben und Lehrsätze aus der Analysis*, vol. 2 (1925), pp. 142, 357, problem 156.

[4] R. Salem, "Sets of uniqueness and sets of multiplicity," *Transactions of the American Mathematical Society*, vol. 54 (1943), pp. 218-228.

[5] ———, "A remarkable class of algebraic integers. Proof of a conjecture of Vijayaraghavan," *Duke Mathematical Journal*, vol. 11 (1944), pp. 103-108.

[6] ———, "Power series with integral coefficients," *ibid.*, vol. 12 (1945), pp. 153-172.

[7] C. L. Siegel, "Algebraic integers whose conjugates lie in the unit circle," *ibid.*, vol. 11 (1944), pp. 597-602.

DYNAMICAL SYSTEMS WITH INDETERMINACY.*[1]

By WILFRED KAPLAN.

1. INTRODUCTION.

1. 1. Classical dynamical systems. The classical theory of dynamical systems concerns the properties of the solutions of differential equations

$$dx_i/dt = f_i(x_1, \cdots, x_n) \qquad (i = 1, \cdots, n)$$

(in local coordinates) in an n-dimensional manifold M, the phase space. This theory has been developed to a considerable extent through the efforts of Poincaré, G. D. Birkhoff and others. The results obtained concern methods of obtaining solutions (usually in series form), existence and stability of periodic solutions, nature of singular points, existence of wandering motions, recurrent motions and other special types of solutions, and ergodic theory.

In general one is impressed with the complexity of the problem. For $n = 2$ a complete analysis of all possibilities appears achievable, but for $n = 3$ or more such a program appears hopeless.

1. 2. Introduction of indeterminacy. In the present paper the statement of the problem is modified as follows. Instead of studying the exact solutions of the above differential equations, one considers the ϵ-solutions, solutions $x_i = x_i(t)$ for which the velocity vector dx_i/dt differs from the prescribed vector f_i by 'less than ϵ' (a condition to be made precise below). In effect one thus replaces the given direction at each point by a 'cone of directions,' so that the direction of the solution at each point has a certain degree of indeterminacy.

Such a modification is not unfamiliar, a similar approach being used in the theory of random processes. Here, however, the probability aspect will not be emphasized, the answers to all questions being given in the form of "yes" or "no" (possibility or impossibility).

The reasons for the introduction of indeterminacy are as follows. It is felt that this is a more realistic formulation of the physical problem for

---

* Received November 9, 1949.

[1] Presented to the American Mathematical Society on April 27, 1946 and September 9, 1948. The present paper is a result of research initiated under the sponsorship of the Office of Naval Research and completed while the author was a Guggenheim Fellow.

which the theory of dynamical systems has been devised. In every concrete physical problem a certain amount of approximation is needed before the problem can be reduced to precise mathematical form. In many cases the approximations used can be estimated, at least roughly, so that one has some knowledge concerning the "ϵ" involved. Thus there appears to be a natural physical basis for the present approach. But, as will be seen below, the introduction of indeterminacy has the further advantage of greatly simplifying the mathematical description of the family of solutions.

1. 3. Summary of results. One introduces a partial order in M by the definition: $x < y$ if there is an ϵ-solution from x to y with increasing t. It may then happen that $x < y < x$ for some x and y, so that there is a periodic ϵ-solution through x and y. In such a case, or if x coincides with y, one writes: $x \equiv y$, and thereby defines an equivalence relation which decomposes M into equivalence classes, termed "states." Each state consists either of one element, in which case the state is termed *transitory*, or of the points of an open subset of M, in which case the state is termed *stationary*. The state containing x is denoted by $\pi(x)$, the set of all y such that $y < x$ is denoted by $\alpha(x)$. A stationary state $\pi(x)$ is called *ω-stable* if $x < y$ implies $y \,\varepsilon\, \pi(x)$, *α-stable* if $y < x$ implies $y \,\varepsilon\, \pi(x)$. It is shown that, if M is compact, there are in all a finite number of ω-stable stationary states $\pi(z_1), \cdots, \pi(z_N)$ and every x in M satisfies an inequality: $x < z_j$ for at least one z_j; a similar statement applies with regard to α-stability. The sets $F_j = \alpha(z_j)$ are open and cover M. The nerve of this covering is a complex Φ which describes the qualitative structure of the family of ϵ-solutions. It is shown that the complex Φ is 'in general' insensitive to small changes in the allowed error ϵ. It is further shown that, except for certain critical choices of ϵ, the complex Φ can be computed from an analysis of the given vector field at a finite number of points of M. This last result is the basis of a numerical method for obtaining the ω-stable stationary states and the complex Φ. This is illustrated in the case of the Van der Pol equation.

2. FLOWS IN PARTIALLY ORDERED SETS.

2. 1. Partial order and flow. Let M be a set of elements: $x, y, \cdots$ with order relation $<$ such that $x < y$ and $y < z$ imply $x < z$. Then M will be termed *partially ordered*. The set of all y such that $y < x$ will be denoted by $\alpha(x)$; the set of all y such that $x < y$ will be denoted by $\omega(x)$. If, for each x in M, the set $\alpha(x)$ is non-void, then the partial order will be termed α-extensive; if, for each x in M, the set $\omega(x)$ is non-void, then the partial order will be termed *ω-extensive*; if both $\alpha(x)$ and $\omega(x)$ are non-void for all x in M, then the order will be termed *extensive*. If the relation $x < x$ holds for no x in M, then the partial order will be termed *proper*.

The partial order can be thought of as representing a flow with indeterminacy in M. Thus the relation $x < y$ can be interpreted as meaning: it is possible to move from x to y with increasing time t. In a classical dynamical system the relation $x < y$ would mean simply: the trajectory $z = \phi(t)$ passing through x when $t = 0$ passes through y for some positive t; here the indeterminacy is degenerate: if $x < y$, then it is not merely possible but inevitable that the trajectory through x reaches y at a later time. The indeterminacy can arise, as will be discussed in Section **3** below, through replacing the exact differential equations of a classical dynamical system by inequalities, reflecting incomplete knowledge of the system.

A relation: $x < x$ is simply a statement that there is a closed (periodic) orbit through x. There is thus no reason for excluding this possibility for a general flow. Accordingly, the partial orders consider here are in general *improper*.

2. 2. The states $\pi(x)$. Let M denote a fixed partially ordered set. For each x in M the set $\pi(x)$ is defined by the equation

$$\pi(x) = \{x\} \cup (\alpha(x) \wedge \omega(x)).$$

Thus $\pi(x)$ consists of x plus all y such that $x < y$ and $y < x$, hence of x plus all elements on closed trajectories through x.

If $\alpha(x) \wedge \omega(x) = 0$, so that $\pi(x)$ reduces to $\{x\}$, then the element x will be termed *transitory*. If $\alpha(x) \wedge \omega(x) \neq 0$, then x will be called *stationary*. If x is stationary, then $x \in \alpha(x) \wedge \omega(x)$. For $y \in \alpha(x) \wedge \omega(x)$ implies $x < y$ and $y < x$, so that $x < x$. Hence in this case $\pi(x) = \alpha(x) \wedge \omega(x) \supseteq \{x\}$. The same reasoning shows that, if x is stationary, then every phase y in $\pi(x)$ is stationary.

Each set $\pi(x)$ will be called a *state*, and will be termed a *transitory state* or a *stationary state* according to whether it consists of one transitory element x or of one or more stationary elements $x, y, \cdots$.

2. 3. Ordering of the states. If $\pi(x) \wedge \pi(y) \neq 0$, then $\pi(x) = \pi(y)$. For if $z \in \pi(x)$ and $w \in \pi(x) \wedge \pi(y)$, then $z \leqq x \leqq w \leqq y$ and $y \leqq w \leqq x \leqq z$, so that $z \in \pi(y)$. Thus $\pi(x) \subseteq \pi(y)$ and similarly $\pi(y) \subseteq \pi(x)$, so that $\pi(x) = \pi(y)$.

It follows that the states $\pi(x)$ form a decomposition of M into disjoint subsets. The notation $x \equiv y$ will be used for the corresponding equivalence relation: i.e., $x \equiv y$ if $\pi(x) = \pi(y)$.

If $x \equiv y$, $z \equiv w$, and $x < z$, then $y < w$. For $y \leqq x < z \leqq w$. Hence one can define: $\pi(x) < \pi(y)$ if $\pi(x) \neq \pi(y)$ and $z < w$ for some $z \equiv x$ and some $w \equiv y$. This ordering satisfies the transitive law, so that the states are now partially ordered. This ordering is proper, for $\pi(x) < \pi(x)$ is ruled out by definition.

A stationary state $\pi(x)$ will be called *ω-stable* if $\pi(x)$ is maximal, i. e., if there is no $\pi(y)$ such that $\pi(y) > \pi(x)$. Similarly, $\pi(x)$ is *α-stable* if $\pi(x)$ is minimal.

The given partially ordered set M will be said to have *finite ω-stability type* if the following two conditions hold:

(a) the number of ω-stable states is finite;

(b) for every state $\pi(x)$ there exists an ω-stable state $\pi(z)$ such that $\pi(x) \leqq \pi(z)$.

The notion of *finite α-stability type* is defined in analogous fashion.

2. 4. Stability complexes. Let M have finite ω-stability type and let $\pi(x_1), \cdots, \pi(x_N)$ be the ω-stable states. By condition (b) the sets $F_1 = \alpha(x_1), \cdots, F_N = \alpha(x_N)$ cover M. The nerve of the covering is a finite simplicial complex Φ, the *ω-stability complex.* Thus the N vertices of Φ are the sets $F_1, \cdots, F_N$, the edges are those pairs (F_j, F_k) for which $j \neq k$ and $F_j \wedge F_k \neq 0$, the 2-simplexes are those triples (F_h, F_j, F_k) for which $h \neq j$, $j \neq k$, $k \neq h$ and $F_h \wedge F_j \wedge F_k \neq 0$, etc. In the same way, if M has finite α-stability type, the covering of M by sets G_k forms a finite complex Γ, the *α-stability complex.*

Each simplex of Φ corresponds to a subset of M with given uncertainty about the ultimate future. Thus the 2-simplex (F_1, F_2, F_3) corresponds to the subset $F_1 \wedge F_2 \wedge F_3$ of those elements from which the states $\pi(x_1)$, $\pi(x_2)$, $\pi(x_3)$ are possible future states. In dynamical language, these elements are the initial conditions from which the ω-stable states $\pi(x_1)$, $\pi(x_2)$, $\pi(x_3)$ can be reached. Similar remarks apply to Γ with regard to the past.

3. DYNAMICAL SYSTEMS WITH INDETERMINACY.

3. 1. The phase space *M*. It will be assumed that M is a compact n-dimensional manifold of class $C^{(3)}$. It will further be assumed that M has a Riemannian metric: $ds^2 = a_{ij} dx^i dx^j$ in terms of local coordinates

$(x^1, \cdots, x^n)$, where the a_{ij} are of class C'. Hence M becomes a metric space with metric $\rho(x, y)$ equal to the g. l. b. of the lengths of all paths from x to y in M. Finally, it will be assumed that a fixed *dynamical system* is given in M: i. e., a vector field $v = v(x)$ whose contravariant components $v^i(x^1, \cdots, x^n)$ are of class C'. The differential equations

$$dx^i/dt = v^i(x^1, \cdots, x^n) \qquad (i = 1, \cdots, n) \tag{1}$$

then determine the trajectories of the given dynamical system. The points of M will be referred to as *phases*; M will be termed the *phase space.*

Associated with M is the fibre space V of all vectors u in M. If local coordinates $(x^1, \cdots, x^n)$ are chosen in M, then local coordinates $(x^1, \cdots, x^n; u^1, \cdots, u^n)$ are determined in V; for this reason the points of V will be denoted by $(x; u)$. The vectors u at each point form a normed linear space with norm $\| u \|$ equal to $(a_{ij}u^iu^j)^{\frac{1}{2}}$. The subset of V of all vectors with non-zero norm will be denoted by V'.

The space V, with the above coordinate systems, is a differentiable manifold and, as a normal separable space, is metrisable. The metric will be denoted by $\sigma((x; u), (y; w))$.

3.2. Indeterminacy functions. By an *indeterminacy function* in M will be meant a real-valued function $\epsilon(x; u)$ defined in V and continuous in V' and satisfying the conditions:

$$\epsilon(x; u) > 0; \quad \epsilon(x; u) = \epsilon(x; ku), k > 0; \quad \epsilon(x; 0) = 1. \tag{2}$$

Thus ϵ is a function of direction only. For many purposes it will be convenient to use the related function $E(x; u)$ defined by the equation:

$$E(x; u) = \| u \| / \epsilon(x; u). \tag{3}$$

Thus $E(x; u)$ is continuous in V and $E(x; ku) = kE(x; u)$ for $k > 0$.

The indeterminacy function $\epsilon(x; u)$ will be said to be *convex* if the set of all vectors u at x such that $\| u \| < \epsilon(x; u)$ is convex. This is equivalent to the condition that the set of u such that $E(x; u) < 1$ be convex or, because of the homogeneity of E, to the condition that

$$E(x; \lambda u_1 + (1 - \lambda) u_2) \leqq \lambda E(x; u_1) + (1 - \lambda) E(x; u_2), \quad 0 \leqq \lambda \leqq 1. \tag{4}$$

Corresponding to the indeterminacy function $\epsilon(x; u)$ there is an *indeterminacy neighborhood* $U(x; v; \epsilon)$ of each vector v (at x) of the given dynamical system: namely, the set of all vectors w at x that $E(x; w - v) < 1$.

By an *allowed solution relative to* $\epsilon(x; u)$ or, more briefly, by an *ϵ-solution,* will be meant a path $x = x(t)$, defined over some t-interval (perhaps infinite), which is piecewise smooth and is such that at each point x of the path the vector dx/dt satisfies:

$$(5) \qquad \| dx/dt - v(x) \| < \epsilon(x; dx/dt - v(x))$$

(for either choice of dx/dt at corners) or, in terms of the E-function:

$$(5') \qquad E(x; dx/dt - v(x)) < 1.$$

Thus the velocity vector dx/dt is required to lie within the indeterminacy neighborhood $U(x; v(x); \epsilon)$ at each x.

The classical existence theorem for differential equations guarantees existence of a unique solution of (1) through each phase x. Here one can state that *through each phase* x_0 *and for each vector* w_0 *in* $U(x_0; v(x_0); \epsilon)$ *there is an* ϵ*-solution (not unique)* $x = x(t)$ *for which the velocity vector at* x_0 *is* w_0. For, in local coordinates, the straight line $x^i = x_0{}^i + w_0{}^i t$, $|t| < t_1$, will be such a solution, for t_1 sufficiently small. This follows from the fact that $E(x_0 + w_0 t; w_0 - v(x_0 + w_0 t))$ is continuous in t at $t = 0$, is < 1 for $t = 0$ and hence remains < 1 for $|t|$ sufficiently small. An infinity of such solutions can clearly be provided. In general, all the ϵ-solutions through x_0 form a configuration like a spray from a nozzle.

3. 3. Partial order in *M*. A partial order in M is now introduced by the definition: $x < y$ if there is an ϵ-solution passing through x and y for values t_1 and t_2 respectively of t, with $t_1 < t_2$. This clearly obeys the transitive law, while the possibility of periodic solutions makes the order in general improper. The ordering is an extensive one, by virtue of the existence theorem proved in **3. 2**. The dependence of the order on the choice of $\epsilon(x; u)$ will be indicated, where necessary, by the notation: $x < y\ [\epsilon]$. For the present a fixed choice of $\epsilon(x; u)$ will be assumed.

THEOREM 1. *For each* x_0 *in* M*, the sets* $\alpha(x_0)$ *and* $\omega(x_0)$ *are open.*

Proof. Let $y \in \omega(x_0)$, i. e., $x_0 < y$, so that there is an ϵ-solution $x = x(t)$, $t_1 \leqq t \leqq t_2$, with $x(t_1) = x_0$, $x(t_2) = y$, $t_1 < t_2$. Choose local coordinates at y and let t_3 be chosen so that $t_1 < t_3 < t_2$ and so that $x(t)$ lies in the chosen neighborhod of y for $t_3 \leqq t \leqq t_2$. For each constant vector u such that $u^i u^i = 1$ and for λ_1 sufficiently small, the paths

$$x^i = x^i(t) + \lambda u^i(t - t_3), \qquad t_3 \leqq t \leqq t_2, \qquad 0 \leqq \lambda < \lambda_1,$$

then lie within the chosen neighborhood of y and are furthermore ϵ-solutions. This follows at once, since

$$E(x(t)+\lambda u(t-t_3)\,;\,dx/dt+\lambda u-v(x+\lambda u(t-t_3)))$$

is continuous in t and λ and is less than 1 for $\lambda=0$. The paths all pass through $z=x(t_3)$ for $t=t_3$, and for $t=t_2$ fill out a spherical neighborhood of radius $\lambda_1(t_2-t_3)$ about y. Each phase x of this spherical neighborhood satisfies $z<x$ and, by the transitive law, $x_0<x$. Hence $\omega(x_0)$ is open and, in the same way, $\alpha(x_0)$ is open.

COROLLARY. *If x is a stationary phase, then $\pi(x)$ is open.*

For here $\pi(x)=\alpha(x)\triangle\omega(x)$.

THEOREM 2. *M, as partially ordered above, has finite ω-stability type and finite α-stability type.*

Proof. This is an immediate consequence of compactness. For each x in M one can choose $y>x$, since M is extensively ordered. Hence all the sets $\alpha(x)$ together cover M. Hence there exist a finite number of these: $\alpha(x_1),\cdots,\alpha(x_N)$ which together cover M. By removal of some of these sets, if necessary, one can reduce this covering to a *minimal* one, i. e., one such that removal of any one set would destroy the property of covering M. It will be assumed that this has been done, so that $(\alpha(x_1),\cdots,\alpha(x_N))$ is a minimal covering. It now follows that each x_j is stationary and that each $\pi(x_j)$ is ω-stable, for $j=1,\cdots,N$. For let $y\,\varepsilon\,\omega(x_j)$, $y\neq x_j$. Then $y<x_k$ for some k and hence $x_j<x_k$. This implies that $j=k$, since the covering is minimal, and hence $x_j<x_j$, so that x_j is stationary. The same reasoning shows that $y>x_j$ implies that $y\equiv x_j$, i. e., that $\pi(x_j)$ is ω-stable. There can be no further ω-stable states, since the $\alpha(x_j)$ form a covering. Hence there are a finite number, N, of ω-stable states $\pi(x_j)$ ($j=1,\cdots,N$) and the sets $F_j=\alpha(x_j)$ cover M. In the same manner, one concludes that there are a finite number of α-stable stationary states $\pi(y_k)$ ($k=1,\cdots,P$) and the sets $G_k=\omega(y_k)$ cover M. Thus M has finite stability types.

One can accordingly define the complex $\Phi=\Phi(\epsilon)$ as the nerve of the covering by the F_j and the complex $\Gamma=\Gamma(\epsilon)$ as the nerve of the covering by the G_k. The following theory will be given for Φ alone, the theory for Γ being parallel.

4. VARIATION OF THE INDETERMINACY FUNCTION.

4.1. Ordering of indeterminacy functions. The fibre space V^* of unit vectors in M is a compact space, and one can define the norm $\|f\|$ of a continuous real-valued function in V^* as l. u. b. $|f|$ over V^*. One can define the distance between two such functions f_1, f_2 as $\| f_1 - f_2 \|$, and thereby obtains a metric space. Because of (2), the indeterminacy functions $\epsilon(x;u)$ depend on direction only, hence are determined completely by their values on V^*. The norm $\| \epsilon \|$ of $\epsilon(x;u)$ will mean the norm of $\epsilon(x;u)$ restricted to V^*, and the distance $\| \epsilon_1 - \epsilon_2 \|$ is defined in the same way. It follows that the indeterminacy functions also form a metric space, denoted by $\mathcal{E}$. This space will be partially ordered by the definition: $\epsilon_1 \ll \epsilon_2$ if $\epsilon_1(x;u) < \epsilon_2(x;u)$ for each element $(x;u)$ of V'.

For M itself one then concludes that, if $x < y\ [\epsilon_1]$ and $\epsilon_1 \ll \epsilon_2$, then $x < y\ [\epsilon_2]$.

The complex Φ above is determined by a covering of M. For brevity, Φ will also be identified with the covering. Thus the statement that Φ_1 is a refinement of Φ_2 means that each set (vertex) of Φ_1 is a subset of a set of the covering Φ_2. It should be noted that the covering Φ for given ϵ is a minimal one; this follows from the proof of Theorem 2, or from the definition of finite ω-stability type.

THEOREM 3. *Let ϵ_1 and ϵ_2 be indeterminacy functions and let $\Phi_1 = \Phi(\epsilon_1)$ and $\Phi_2 = \Phi(\epsilon_2)$ be the corresponding ω-stability complexes. If $\epsilon_1 \ll \epsilon_2$, then Φ_1 is a refinement of Φ_2.*

Proof. Let $\pi(x_1), \cdots, \pi(x_{N_1})$ be the ω-stable stationary states relative to ϵ_1; let $\pi(y_1), \cdots, \pi(y_{N_2})$ be the ω-stable states relative to ϵ_2. Let $F_{1j} = \alpha(x_j)$ (relative to ϵ_1), let $F_{2k} = \alpha(y_k)$ (relative to ϵ_2) ($j = 1, \cdots, N_1$, $k = 1, \cdots, N_2$). Then for each x_j there is some y_k such that $x_j \leqq y_k\ [\epsilon_2]$. For each x in F_{1j} one has $x < x_j\ [\epsilon_1]$, hence, since $\epsilon_1 \ll \epsilon_2$, $x < x_j\ [\epsilon_2]$, and hence $x < y_k\ [\epsilon_2]$. Thus $F_{1j} \subseteqq F_{2k}$. Hence Φ_1 is a refinement of Φ_2.

4.2. Critical and non-critical indeterminacy functions. *Definition.* The covering Φ_1 of M by sets F_{1j} $(j = 1, \cdots, N_1)$ is a *similar refinement* of the covering Φ_2 by sets F_{2k} $(k = 1, \cdots, N_2)$ if $N_1 = N_2 = N$ and, for proper numbering, one has $F_{1j} \subseteqq F_{2j}$ $(j = 1, \cdots, N)$ and $F_{1j_1} \wedge F_{1j_2} \wedge \cdots \wedge F_{1j_s} \neq 0$ if and only if $F_{2j_1} \wedge F_{2j_2} \wedge \cdots \wedge F_{2j_s} \neq 0$. In this case Φ_1 and Φ_2 are isomorphic complexes.

THEOREM 4. *For each indeterminacy function $\epsilon = \epsilon(x;u)$ there exists a number η, $\eta > 0$, such that, if ϵ_1 and ϵ_2 are indeterminacy functions, with $\epsilon_1 \ll \epsilon_2 \ll \epsilon$ and $\| \epsilon_1 - \epsilon \| < \eta$, $\| \epsilon_2 - \epsilon \| < \eta$ then $\Phi_1 = \Phi(\epsilon_1)$ is a similar refinement of $\Phi_2 = \Phi(\epsilon_2)$.*

Proof. For each λ between 0 and 1 let $N_0(\lambda)$ equal the number of sets (vertices) in the covering $\Phi(\lambda\epsilon)$. As λ increases, $N_0(\lambda)$ is non-increasing.

For, by Theorem 3, each set of $\Phi(\lambda'\epsilon)$ is contained in a set of $\Phi(\lambda''\epsilon)$ for $\lambda' < \lambda''$ and every set of $\Phi(\lambda''\epsilon)$ must appear in this correspondence, since the coverings are minimal. Hence, for λ_0 sufficiently large and < 1, the number $N_0(\lambda)$ is constant for $\lambda_0 \leqq \lambda < 1$, and the refinement correspondence between $\Phi(\lambda'\epsilon)$ and $\Phi(\lambda''\epsilon)$ is one-to-one for $\lambda_0 \leqq \lambda' < \lambda'' < 1$. On the other hand, for $\lambda_0 \leqq \lambda < 1$, the number $N_1(\lambda)$ of edges in $\Phi(\lambda\epsilon)$ must be non-decreasing. For, if two sets in $\Phi(\lambda'\epsilon)$ intersect, the corresponding pair in $\Phi(\lambda''\epsilon)$ also intersect. A similar statement applies to the higher-dimensional faces. Hence, for λ_0 sufficiently large, the number $N_p(\lambda)$ of p-dimensional faces of $\Phi(\lambda\epsilon)$ is constant for all p and for $\lambda_0 \leqq \lambda < 1$. The same reasoning shows that, for this choice of λ_0, the refinement correspondence between $\Phi(\lambda'\epsilon)$ and $\Phi(\lambda''\epsilon)$, for $\lambda_0 \leqq \lambda' < \lambda'' < 1$ satisfies the definition above for a similar refinement.

Now let $\eta = \text{g. l. b. } (\epsilon(x;u) - \lambda_0\epsilon(x;u)) = (1-\lambda_0) \text{ g. l. b. } \epsilon(x;u)$, where the g. l. b. is over V'. Then $\eta > 0$ and $\epsilon_2 \ll \epsilon$, $\| \epsilon_2 - \epsilon \| < \eta$ imply $\lambda_0\epsilon \ll \epsilon_2 \ll \lambda'\epsilon$ for λ' sufficiently large and less than 1. This implies that $\Phi(\epsilon_2)$ is a refinement of $\Phi(\lambda'\epsilon)$, while $\Phi(\lambda_0\epsilon)$ is a refinement of $\Phi(\epsilon_2)$. Since $\Phi(\lambda_0\epsilon)$ is a similar refinement of $\Phi(\lambda'\epsilon)$, it follows that $\Phi(\epsilon_2)$ is a similar refinement of $\Phi(\lambda'\epsilon)$. If further $\epsilon_1 \ll \epsilon_2 \ll \epsilon$ and $\| \epsilon_1 - \epsilon \| < \eta$, then $\Phi(\epsilon_1)$ is also a similar refinement of $\Phi(\lambda'\epsilon)$. Since $\Phi(\epsilon_1)$ is also a refinement of $\Phi(\epsilon_2)$, this last refinement must be a similar one. Hence the theorem follows.

Definition. An indeterminacy function $\epsilon(x;u)$ is termed *non-critical* if there exists a number $\zeta > 0$, such that $\| \epsilon_1 - \epsilon \| < \zeta$, $\| \epsilon_2 - \epsilon \| < \zeta$, $\epsilon_1 \ll \epsilon_2$ imply that $\Phi(\epsilon_1)$ is a similar refinement of $\Phi(\epsilon_2)$. If the condition fails, ϵ is called *critical.* It follows from the definition that the non-critical functions form an open set in $\mathcal{E}$.

THEOREM 5. *The critical indeterminacy functions form a nowhere dense set in* $\mathcal{E}$.

Proof. Let ϵ be critical and choose $\eta = \eta(\epsilon)$ as in the preceding theorem. For given $\delta > 0$, one can then choose $\epsilon_3 = \lambda\epsilon$ such that $\epsilon_3 \ll \epsilon$, $\| \epsilon - \epsilon_3 \| < \eta/2$ and $\| \epsilon - \epsilon_3 \| < \delta/2$. Choose $\zeta = \frac{1}{2} \text{ g. l. b. } (\epsilon - \epsilon_3)$ over V'. Then $\zeta > 0$ and $\| \epsilon_j - \epsilon_3 \| < \zeta$ imply $\epsilon_j \ll \epsilon$ and $\| \epsilon_j - \epsilon \| < \eta$ for $j = 1, 2$. Hence, by the preceding theorem, if $\epsilon_1 \ll \epsilon_2$, Φ_1 is a similar refinement of Φ_2. Thus ϵ_3 is non-critical. Since the non-critical functions form an open set, it follows that every neighborhood of ϵ contains an open set of non-critical functions. Thus the critical functions form a nowhere dense set.

5. APPROXIMATION BY FLOW ON A NET.

5.1. Flows in M and in M_0. The purpose of this section is to show that, when ϵ is non-critical and convex, the complex Φ can be found by a finite computation, based on a study of the field $v(x)$ on a finite subset M_0 of the space M. One can first *project* the flow in M onto a flow on M_0 by requiring that $x < y$ hold in M_0 precisely when it holds in M. This clearly determines a partial order in M_0. If this partial order is ω-extensive, one can then apply all the theory of Section 2 above to obtain stationary states, transitory states and (the ω-stability type being necessarily finite because M_0 is finite) the ω-stability complex. Corresponding to a complex Φ_1 in M one thus obtains a complex $\Phi_1{}^0$ in M_0. It is further clear that, if M_0 is sufficiently dense in M, the ordering in M_0 will be ω-extensive and Φ_1 and $\Phi_1{}^0$ will be isomorphic. "Sufficiently dense" here means that there must be at least one point of M_0 in each ω-stable state $\pi(x_j)$ and that, whenever an intersection $F_{j_1} \wedge F_{j_2} \wedge \cdots \wedge F_{j_s}$ is non-void, then the intersection contains a point of M_0. Since the sets concerned here are open, a $\delta > 0$ exists such that, if M_0 is δ-dense in M (every point of M within distance δ of a point of M_0), then these conditions hold, and Φ_1 is isomorphic to $\Phi_1{}^0$.

The process just described does not actually reduce the determination of Φ_1 to a computation on M_0 alone, since the validity of each inequality must be investigated in M. Accordingly, a special process must be devised, involving M_0 alone. The following is a description of the process.

A finite set which is δ_1-dense in M will be called a δ_1-*net*. It is known in differential geometry (cf. H. Seifert and W. Threlfall, *Variationsrechnung im Grossen*, Berlin (1938), p. 49) that there exists a positive number d such that each two points of M within distance d can be joined by a unique minimizing geodesic. Let δ_2 satisfy: $0 < \delta_2 < d$. If M_0 is a δ_1-net and each two points of M_0 within distance δ_2 have been joined by the corresponding geodesic, then M_0 becomes a $\delta_1\delta_2$-*network*. The points of M_0 will be numbered as z_μ $(\mu = 1, \cdots, m)$ and the geodesic arcs will be denoted by $C_{\mu\nu}$ for appropriate values of μ and ν. The given vector v at z_μ will be denoted by v_μ and the unit tangent vector to $C_{\mu\nu}$ at z_μ in the direction of z_ν will be denoted by $u_{\mu\nu}$.

Let $\epsilon^0(z_\mu; u)$ now be an indeterminacy function in M_0, i. e., a function satisfying (2) and continuous in u for $u \neq 0$ and z_μ in M_0. Let $C_{\mu\nu}$ be an arc of the $\delta_1\delta_2$-network. If, for some $k > 0$,

$$\| k u_{\mu\nu} - v_\mu \| < \epsilon^0(z_\mu; k u_{\mu\nu} - v_\mu), \tag{6}$$

then z_ν will be termed a *successor of* z_μ. If $z_{\mu_1}, z_{\mu_2}, \cdots, z_{\mu_p}$ are points of M_0 such that z_{μ_k} is a successor of $z_{\mu_{k-1}}$ for $k = 2, \cdots, p$, then the inequality: $z_{\mu_1} \prec z_{\mu_p}$ will be said to hold and the succession of arcs $C_{\mu_1\mu_2}, C_{\mu_2\mu_3}, \cdots, C_{\mu_{p-1}\mu_p}$ together form an *allowed* ϵ^0-*trajectory* on the $\delta_1\delta_2$-network. This definition determines a partial order in M_0. If this partial order is ω-extensive, the theory of Section **2** can be applied, the order has finite stability type, and a complex Φ^0 is determined. The main goal of this section is to establish the following theorem.

THEOREM 6. *Let* ϵ_1 *be a non-critical convex indeterminacy function, of class* C' *in* u *for each* x, $u \neq 0$. *Let* M_0 *be a* $\delta_1\delta_2$-*network based on the points* z_μ $(\mu = 1, \cdots, m)$. *Then, for* δ_1, δ_2, *and* δ_1/δ_2 *sufficiently small and* $\epsilon^0 = \epsilon_1(z_\mu; u)$, *the ordering* $z_\mu \prec z_\nu\ [\epsilon^0]$ *is* ω-*extensive and the complex* Φ^0 *is isomorphic to the complex* Φ_1.

5. 2. Two fundamental lemmas. In order to prove the theorem just stated, one first chooses constants λ_2 and λ_3 so that $\epsilon_2 = \lambda_2\epsilon_1 << \epsilon_1 << \epsilon_3 = \lambda_3\epsilon_1$ and $\| \epsilon_2 - \epsilon_1 \| < \zeta(\epsilon_1)$, $\| \epsilon_3 - \epsilon_1 \| < \zeta(\epsilon_1)$. It follows that the complexes $\Phi_j = \Phi(\epsilon_j)$ $(j = 1, 2, 3)$ are isomorphic, the covering Φ_2 being a similar refinement of Φ_1, the covering Φ_1 being a similar refinement of Φ_3. Let M_0 be a $\delta_1\delta_2$-network, with δ_1 so small that the projections $\Phi_2{}^0$ and $\Phi_3{}^0$ are defined and are isomorphic to Φ_2 and Φ_3 respectively; hence $\Phi_2{}^0$ and $\Phi_3{}^0$ are isomorphic, and $\Phi_2{}^0$ is a similar refinement of $\Phi_3{}^0$.

LEMMA 1. *For* δ_1, δ_2 *and* δ_1/δ_2 *sufficiently small,* $z_\mu < z_\nu\ [\epsilon_2]$ *implies* $z_\mu \prec z_\nu\ [\epsilon^0]$.

LEMMA 2. *For* δ_1 *and* δ_2 *sufficiently small,* $z_\mu \prec z_\nu\ [\epsilon^0]$ *implies* $z_\mu < z_\nu\ [\epsilon_3]$.

The proofs of these lemmas will be given in the following sections. Here it will be shown that they imply Theorem 6.

By the above construction, δ_1 is chosen so small that the ordering $z_\mu < z_\nu\ [\epsilon_2]$ is ω-extensive on M_0. Lemma 1 then implies that the ordering $z_\mu \prec z_\nu\ [\epsilon^0]$ is ω-extensive on M_0, so that a complex Φ^0 is obtained. By virtue of Lemma 1, the covering $\Phi_2{}^0$ is a refinement of the covering Φ^0. By virtue of Lemma 2, the covering Φ^0 is a refinement of the covering $\Phi_3{}^0$. Since $\Phi_2{}^0$ is a similar refinement of $\Phi_3{}^0$, it follows that $\Phi_2{}^0$ is a similar refinement of Φ^0 and that Φ^0 is a similar refinement of $\Phi_3{}^0$. Hence $\Phi_2{}^0$, Φ^0, $\Phi_3{}^0$ are all isomorphic. Since $\Phi_2{}^0$ is isomorphic to Φ_2, and Φ_2 to Φ_1, the theorem is thus established.

5. 3. Auxiliary lemmas. Before proceeding to the proofs of Lemmas 1 and 2, it will be useful to carry out several preliminary steps which will simplify the uniformity arguments to follow.

At each phase x_0 of M a coordinate system can be chosen for a neighborhood of x_0 and an $r > 0$ can then be chosen so that the Euclidean sphere S: $\Sigma(x^i - x_0{}^i)^2 \leqq r^2$, in these coordinates, lies within the coordinate neighborhood. Further, a number $R > 0$ can then be chosen so that the neighborhood $\rho(x_0, x) < R$ lies in S. The collection of all neighborhoods $W: \rho(x_0, x) < \frac{1}{2}R$ covers M. Hence a finite number of these: $W_q: \rho(y_q, x) < \frac{1}{2}R_q$ $(q = 1, \cdots, Q)$ covers M. With each W_q there is associated a coordinate system and a Euclidean sphere S_q in these coordinates such that the neighborhood $\rho(y_q, x) < R_q$ lies in S_q. In the following proof only these coordinate systems will be used.

Let $R_0 = \min(R_1, \cdots, R_Q)$. If $\rho(z_\mu, z_\nu) < d$ and $\rho(z_\mu, z_\nu) < R_0/4$, then $\rho(z_\mu, y_q) < \frac{1}{2}R_q$ and $\rho(z_\nu, y_q) < 3R_q/4$ for some q. Hence both the geodesic $C_{\mu\nu}$ and the Euclidean straight line, in terms of the chosen coordinates at y_q, lie within S_q.

Throughout the following it will often be convenient to use the E-function (**3. 2** above) corresponding to a given indeterminacy function $\epsilon(x; u)$. In each case the matching of subscripts or superscripts will indicate the corresponding pairs of functions.

LEMMA 3. *Let ϵ_1 and ϵ_2 be indeterminacy functions in M such that $\epsilon_1 \ll \epsilon_2$. Then there exists a constant h such that, if x_1 and x_2 are in S_q and $(x_1; u_1)$, $(x_2; u_2)$ are such that*

$$E_1(x_1; u_1 - v(x_1)) < 1, \quad | x_1{}^i - x_2{}^i | < h, \quad | u_1{}^i - u_2{}^i | < h$$

in the corresponding coordinates, then $E_2(x_2; u_2 - v(x_2)) < 1$.

Proof. Let I_j $(j = 1, 2)$ denote the subset of V consisting of all elements $(x; w)$ such that $E_j(x; w - v(x)) < 1$. Then, since the E_j and v are continuous, I_j is open. Further, because of the homogeneity of the E_j, the closure $\bar{I}_1$ of I_1 is compact and lies in I_2. Hence the distance σ_0 between I_1 and the boundary of I_2 is positive and, if $(x_1; u_1)$ is in I_1 and $\sigma((x_1; u_1), (x_2; u_2)) < \sigma_0$, then $(x_2; u_2)$ is in I_2. Because of the compactness of $\bar{I}_1$ and the equivalence of the topology in V with the Euclidean topology defined by the coordinate system in S_q, a number h_q can be found such that, if $(x_1; u_1)$ is in I_1 and $| x_1{}^i - x_2{}^i | < h_q$, $| u_1{}^i - u_2{}^i | < h_q$ then $\sigma((x_1; u_1), (x_2; u_2)) < \sigma_0$. If h is now chosen as $\min(h_1, \cdots, h_Q)$, then the conclusion follows.

The geodesics in M are defined by equations of the form:

$$(7)\qquad d^2x^i/ds^2 = f^i(x^1,\cdots,x^n,dx^1/ds,\cdots,dx^n/ds),$$

where the f^i are continuous. The variables dx^i/ds are the components of a unit vector and hence are bounded (in absolute value), in the local coordinates, in S_q by a constant H_q. Thus $H = \max(H_1,\cdots,H_Q)$ is a bound for these components on M. In each set S_q, the point $(x^1,\cdots,x^n, dx^1/ds,\cdots,dx^n/ds)$ ranges over a compact subset of V. Hence the components f^i are uniformly bounded in S_q by a constant F_q and, if $F = \max(F_1,\cdots,F_Q)$, then F serves as a uniform bound for these components on M. Similarly, the coefficients a_{ij} of the fundamental form ds^2 are bounded by a constant A on M.

LEMMA 4. *Let x_1x_2 be a geodesic in M of length s_1 lying completely in a set S_q. Then one has*

$$(8)\quad |x_2{}^i - x_1{}^i| < Fs_1{}^2 + Hs_1,$$
$$|u_2{}^i - u_1{}^i| < Fs_1,\ |u_1{}^i - (x_2{}^i - x_1{}^i)/s_1| < Fs_1{}^2,$$

where $u_j = dx/ds$ at x_j.

Proof. The differential equations (7) above can be integrated once along x_1x_2 to give

$$(9)\qquad dx^i/ds = \int_0^s f^i ds + u_1{}^i,$$

and again to give

$$(10)\qquad x^i = \int_0^s \int_0^s f^i ds ds + u_1{}^i s + x_1{}^i.$$

The inequalities (8) follow at once from (9) and (10) and from the fact that $|f^i| < F$, $|u_j{}^i| < H$.

LEMMA 5. *Let D denote the n-dimensional space of coordinates $(x^1,\cdots, x^n)$ with metric $ds^2 = a_{ij}dx^idx^j$, the a_{ij} being constants. Let $v(x)$ be a vector field defining a dynamical system in D, whereby $v(x)$ is independent of x, so that $v(x) = v(0)$. Let $\epsilon(x;u)$ be a convex indeterminacy function in D, of class C' for $u \neq 0$, which is independent of x, so that $\epsilon(x;u) = \epsilon(u)$. Let y and z be in D and let $y < z\ [\epsilon]$. Let yz be an allowed ϵ-trajectory from y to z, and let τ be the time interval on this trajectory. Then the uniform trajectory:*

$$(11)\qquad x^i = y^i + (z^i - y^i)t/\tau \qquad (i = 1,\cdots,n)$$

from y to z is also an allowed ϵ-trajectory from y to z.

Proof. The equations

$$x^i = y^i + t(v^i + u^i \varepsilon(u)),\ t \geqq 0, \tag{12}$$

where u is a variable unit vector, define a semi-cone in $x^1 \cdots x^n t$-space. By the "inside" of this cone will be meant the domain

$$x^i = y^i + t(v^i + w^i \epsilon(w)),\ t > 0, \tag{13}$$

where w is a variable vector with $\| w \| < 1$. The trajectory yz corresponds to a curve

$$x^i = \phi^i(t), \qquad t = t \qquad (0 \leqq t \leqq \tau) \tag{14}$$

in $x^1 \cdots x^n t$-space, passing through the vertex $(y^1, \cdots, y^n, 0)$ of the cone.

Now the function $E(x; u) = E(u)$ corresponding to $\epsilon(u)$ is positive, homogeneous of degree one, and the sets $E \leqq$ const. are convex. From these two properties one concludes that, for any points x and z $(x \neq 0)$ one has

$$z^i \partial E/\partial x^i \leqq E(z)\,. \tag{15}$$

For if z lies in the tangent plane to the surface $E =$ const. at x, then $(x^i - z^i)\partial E/dx^i = 0$ or, by the Euler theorem, $E(x) = z^i \partial E/\partial x^i$. Since z cannot lie inside the surface $E =$ const., one has $E(x) \leqq E(z)$, so that (15) holds. The homogeneity of $E(z)$ then implies (15) for arbitrary z.

Now let $g(x, t)$ be defined by the equation

$$g(x, t) = E(x - y - vt). \tag{16}$$

Then on the cone (12) one has

$$g(x, t) = \| x - y - vt \| / \{\epsilon(x - y - vt)\} = t\epsilon(u)/\{\epsilon(tu\epsilon(u))\} = t, \tag{17}$$

and, by a similar reasoning, $g(x, t) > t$ for $t > 0$ and (x, t) outside the cone (12). On the path (14) one has $E(dx/dt - v(x)) < 1$. Hence, by (15), at each point on this path $x = y + vt$, or

$$dg/dt = (dx^i/dt - v^i)\partial E/\partial x^i \leqq E(dx/dt - v) < 1. \tag{18}$$

It follows that the path (14) enters the interior of the cone at $t = 0$ and remains inside. Hence the end-point z is inside the cone, and the straight line path (11) is inside the cone. Hence this path can be written in the form (13) with constant w, $\| w \| < 1$. Hence on (11)

$$\| dx/dt - v \| = \|w\| \epsilon(w) < \epsilon(w) = \epsilon(dx/dt - v). \tag{19}$$

Thus (11) is an allowed ϵ-trajectory, and Lemma 5 is established.

Remark. The assumption that $\epsilon(u)$ is of class C' is clearly needed for the proof just given, and this is the basis for the corresponding assumption in Theorem 6. It is however possible to prove, without too much trouble, that a convex indeterminacy function $\epsilon(u)$ can be arbitrarily closely approximated by a convex function of class C'. By means of this approximation, Lemma 5 can be applied to prove Theorem 6 without the assumption that $\epsilon_1(x; u)$ be of class C'. The proof is given in the present form for the sake of simplicity.

LEMMA 6. *Let $\epsilon(x; u)$ be a given indeterminacy function in M. Then there exist constants B_1, B_2, C such that*

$$| dx^i/dt | < B_1, \qquad | dx^i/dt - v^i | < B_2, \qquad s < C\tau, \tag{20}$$

in the coordinates of each set S_q, where $x^i = x^i(t)$, $t_1 \leqq t \leqq t_1 + \tau$, is an ϵ-solution from x_1 to x_2 of length s in S_q.

Proof. The function $\epsilon(x; u)$ is uniformly bounded by a constant ϵ^*. Hence $\| dx/dt - v(x) \| < \epsilon^*$. The set of all $(x; w)$ such that x is in S_q and $\| w \|^2 = a_{ij}w^iw^j \leqq \epsilon^{*2}$ is a compact subset of V. Hence its projection on the w^i-axis is bounded. Hence a constant B_2 can be chosen, independent of q, so that $\| w \|^2 \leqq \epsilon^{*2}$ implies $| w^i | < B_2$. Thus on the ϵ-trajectory $| dx^i/dt - v^i | < B_2$. The vector v has also a maximum norm K_1 on M. Hence there is a constant K_2 (independent of q), such that in each S_q $| v^i | < K_2$. Hence $| dx^i/dt | < K_2 + B_2 = B_1$. The velocity on an ϵ-solution is bounded by $K_1 + \epsilon^* = C$. Hence s/τ, which is simply the average velocity, is bounded by C.

5.4. Proof of Lemma 1. *Remark.* Throughout the following proofs the choices of δ_1 and δ_2 will be continually restricted. In each case the restriction will be a uniform one, i. e., dependent only on the given vector field $v(x)$ and indeterminacy function $\epsilon_1(x; u)$ and not on the particular points or neighborhoods considered.

Let $\epsilon_4 = \lambda_4 \epsilon_1$ and $\epsilon_5 = \lambda_5 \epsilon_1$ be chosen so that $\epsilon_2 \ll \epsilon_4 \ll \epsilon_5 \ll \epsilon_1$. Choose δ_2 less than $R_0/4$.

Suppose now that $z_\mu < z_\nu\ [\epsilon_2]$, so that there is an allowed ϵ_2-solution in M from z_μ to z_ν.

Case A. *The ϵ_2-solution $z_\mu z_\nu$ from z_μ to z_ν lies within the δ_2-neighborhood of z_μ.* One can then choose local coordinates in a sphere S_q as above, with $\rho(z_\mu, y_q) < R_q/2$.

It follows from Lemma 3 that, for δ_2 sufficiently small, one has at each phase x of the trajectory $z_\mu z_\nu$ $E_4(z_\mu; dx/dt - v(x)) < 1$, so that $z_\mu z_\nu$ is an allowed solution relative to the constant (independent of x) indeterminacy function $\epsilon' = \epsilon_4(z_\mu; u)$ in S_q. Thus one has at each point of $z_\mu z_\nu$

$$(21) \qquad \| dx/dt - v(x) \| < \epsilon'(dx/dt - v(x)).$$

Since the a_{ij} are continuous, one can further choose δ_2 so small that at each point of $z_\mu z_\nu$ one has

$$(22) \qquad \| dx/dt - v(x) \|_\mu < \epsilon(dx/dt - v(x)),$$

where $\| w \|_\mu = (a_{ij}(z_\mu) w^i w^j)^{\frac{1}{2}}$ and $\epsilon(w) = \epsilon_5(z_\mu; w)$. For let $\xi_1 =$ g. l. b. $(\epsilon_5(x; u) - \epsilon_4(x; u))$ on V', so that $\xi_1 > 0$. Given $\xi_2 > 0$, then δ^* can be chosen so that $| a_{ij}(x) - a_{ij}(z_\mu) | < \xi_2$ for $\rho(x, z_\mu) < \delta^*$, whereby δ^* can be chosen independent of x and the particular S_q. Hence, for $\delta_2 < \delta^*$,

$$\begin{aligned} \big| \| dx/dt - v \|^2 - \| dx/dt - v \|_\mu^2 \big| \\ &= \big| (a_{ij}(x) - a_{ij}(z_\mu))(dx^i/dt - v^i)(dx^j/dt - v^j) \big| \\ &< \xi_2 \sum_{i,j} | dx^i/dt - v^i | \, | dx^j/dt - v^j | < \xi_2 n^2 B_2^2, \end{aligned}$$

where the last inequality follows from Lemma 6, since $z_\mu z_\nu$ is an ϵ_2-solution. Since in general $\big| |a| - |b| \big| \leqq (|a^2 - b^2|)^{\frac{1}{2}}$, one concludes that

$$\big| \| dx/dt - v \| - \| dx/dt - v \|_\mu \big| < n B_2 \xi_2^{\frac{1}{2}},$$

hence

$$\| dx/dt - v \|_\mu < \epsilon'(dx/dt - v) + n B_2 \xi_2^{\frac{1}{2}},$$

by (21). If now ξ_2 is chosen so that one has $n B_2 \xi_2^{\frac{1}{2}} < \xi_1$ and δ^* is chosen accordingly, then (22) follows.

Thus $z_\mu z_\nu$ is an ϵ-solution in $x^1 \cdots x^n$-space, with constant metric, where $\epsilon(w) = \lambda_5 \epsilon_1(z_\mu; w)$ is of class C' for $w \neq 0$ and convex. Hence Lemma 5 can be applied and one concludes that the uniform trajectory from z_μ to z_ν in S_q is also an ϵ-solution. Thus in particular one has at z_μ (in local coordinates)

$$\|(z_\mu - z_\nu)/\tau - v_\mu\| < \epsilon_5(z_\mu; (z_\mu - z_\nu)/\tau - v_\mu), \tag{23}$$

where τ is the time on $z_\mu z_\nu$.

Now the geodesic $C_{\mu\nu}$ has length $s < \delta_2$ and one has thus, by Lemma 4,

$$|(z_\mu{}^i - z_\nu{}^i)/s - u_{\mu\nu}{}^i| < F\delta_2{}^2. \tag{24}$$

Hence

$$|(z_\mu{}^i - z_\nu{}^i)/\tau - u_{\mu\nu}{}^i(s/\tau)| < F\delta_2{}^2C. \tag{25}$$

For s is at most equal to the length of $z_\mu z_\nu$, so that s/τ is at most equal to C, in accordance with Lemma 6. It now follows from (23) and (25) and Lemma 3 that, for δ_2 sufficiently small one has

$$E_1(z_\mu; u_{\mu\nu}(s/\tau) - v_\mu) < 1. \tag{26}$$

Thus $z_\mu \prec z_\nu$ $[\epsilon^0]$. Thus Lemma 1 is established in Case A.

Case B. *z_ν does not lie within the δ_2-neighborhood of z_μ.* Then proceed along $z_\mu z_\nu$ from $z_\mu = p_0$ until a point p_1 is reached at distance $\delta_2/3$ from z_μ. Proceed from p_1 until a point p_2 at distance $\delta_2/3$ from p_1 is reached, and so on. One will finally reach a point p_k such that there is no point beyond p_k at distance greater than $\delta_2/3$ from p_k. If $\rho(p_k, z_\nu) < \delta_2/6$, then replace the previous choice of p_k by z_ν, so that one now has $\rho(p_{k-1}, p_k) > \delta_2/6$ and further the trajectory $p_{k-1}p_k$ lies within the $2\,\delta_2/3$ neighborhood of p_{k-1}. If $\rho(p_k, z_\nu) \geqq \delta_2/6$, then set $z_\nu = p_{k+1}$. Thus in either case one has a sequence of points on $z_\mu z_\nu : z_\mu = p_0, p_1, \cdot\cdot\cdot, p_h = z_\nu$ such that on each sub-arc $p_j p_{j+1}$ one has the condition: $\rho(p_j, p_{j+1}) \geqq \delta_2/6$, while the arc $p_j p_{j+1}$ lies within the $2\,\delta_2/3$ neighborhood of p_j.

Now each arc $p_j p_{j+1}$ lies within one of the sets S_q. Proceeding as in Case A, one can then show that, for δ_2 small enough, the uniform trajectory from p_j to p_{j+1}, in the local coordinates, is an ϵ_4-solution.

Now there is by assumption a point z_{μ_j} of M_0 within distance δ_1 of each p_j. Let δ_1 be chosen less than $\delta_2/6$ and less than $R_0/4$. Then the set S_q can be chosen to contain not only the trajectory $p_j p_{j+1}$, but also the points z_{μ_j}, $z_{\mu_{j+1}}$ and the two corresponding line segments in the local coordinates. Let τ be the time used on the trajectory $p_j p_{j+1}$. Then one has

$$E_4(x; [p_{j+1} - p_j]/\tau - v(x)) < 1 \tag{27}$$

at each point x of the segment $p_j p_{j+1}$. One has further, by Lemma 4,

$$(28) \qquad | z^i_{\mu_j} - p^i_j | < F\delta_1^2 + H\delta_1 < L\delta_1, \qquad | z^i_{\mu_{j+1}} - p^i_{j+1} | < L\delta_1$$

for a constant L. Furthermore,

$$(29)) \quad \begin{aligned} &| (z^i_{\mu_{j+1}} - z^i_{\mu_j})/\tau - (p_{j+1} - p^i_j)/\tau | \\ &\qquad \leqq \tau^{-1} (| z^i_{\mu_{j+1}} - p^i_{j+1} | + | z^i_{\mu_j} - p^i_j |) \leqq 2L\delta_1/\tau \leqq 2L\delta_1(6C/\delta_2). \end{aligned}$$

For $\rho(p_j, p_{j+1}) \geqq \delta_2/6$, so that, by Lemma 6, the time τ is greater than $\delta_2/6C$. Now let $x = \phi_1(t)$, $t_1 \leqq t \leqq t_1 + \tau$ represent the uniform trajectory from p_j to p_{j+1} and let $x = \phi_2(t)$, $t_1 \leqq t \leqq t_1 + \tau$ represent the uniform trajectory, with the same interval τ, from z_{μ_j} to $z_{\mu_{j+1}}$. Then (28) and (29) imply that there is a constant h such that

$$\sigma((\phi_1(t); d\phi_1/dt), (\phi_2(t); d\phi_2/dt)) < h,$$

and h can be made as small as desired by choosing δ_1 and δ_1/δ_2 sufficiently small. One concludes from Lemma 3 and (27) that, for δ_1 and δ_1/δ_2 sufficiently small, the uniform trajectory $x = \phi_2(t)$ is an ϵ_5-solution. It now follows as in Case A that $z_{\mu_j} \prec z_{\mu_{j+1}} [\epsilon^0]$. Thus, finally, $z_\mu \prec z_\nu [\epsilon^0]$, for z_{μ_1} can be chosen as z_μ, and z_{μ_h} can be chosen as z_ν, and the transitive law can then be applied. Thus Lemma 1 is completely established.

5. 5. Proof of Lemma 2. Suppose first that z_ν is a successor of z_μ. One is then given that, for some $k > 0$,

$$(30) \qquad E_1(z_\mu; ku_{\mu\nu} - v_\mu) < 1.$$

Choose local coordinates in a set S_q containing $C_{\mu\nu}$ as above. Then, by Lemma 4, one has $| u^i - u^i_{\mu\nu} | < F\delta_2$, where u is the tangent vector dx/ds at an arbitrary point x of $C_{\mu\nu}$. Now the values k for which (30) can hold are uniformly bounded on M by a constant k_0. Hence $| ku^i - ku^i_{\mu\nu} | < k_0 F\delta_2$. Since the vector field $v(x)$ is continuous, one can further, by Lemma 4, choose δ_2 so small that $| v^i - v^i_\mu | < \eta_0$, for given $\eta_0 > 0$, where $v = v(x)$ at an arbitrary point x on $C_{\mu\nu}$. Hence $| (ku^i - v^i) - (ku^i_{\mu\nu} - v^i_\mu) | < k_0 F\delta_2 + \eta_0$. Hence, by Lemma 3, if δ_2 is sufficiently small, one has $E_3(x; ku(x) - v(x)) < 1$ at each point x of $C_{\mu\nu}$. It follows that the path $x = x(t)$, from z_μ to z_ν along $C_{\mu\nu}$, such that $dx/dt = ku(x)$ is an ϵ_3-solution, i. e., that $z_\mu < z_\nu [\epsilon_3]$, as was to be shown. If z_ν is not a successor of z_μ, the same conclusion follows by consideration of each link in the chain from z_μ to z_ν. Thus Lemma 2 is established.

6. CASE OF NON-COMPACT M. REMARKS ON APPLICATIONS.

6. 1. Case of non-compact M. It is clear that the compactness of the phase space M plays an important role in the above discussion. If M is not compact, the number of stable stationary states may be infinite, and the approximation by a flow on a finite set is in general not possible.

There are, however, circumstances, of considerable physical interest, under which the above results can be extended to the case of a non-compact manifold M.

First of all, it may happen that there is a compact subset M_1 of M such that each allowed ϵ-trajectory $x = x(t)$ enters and remains in M for t sufficiently large. In this case the discussion of ω-stability can be restricted to M_1 and the previous results can be extended without difficulty.

Secondly, it may be possible to imbed M in a compact space M' in such a manner that the vector field v can be extended continuously to M'. The relationships between the states in M' and those in M will in general not be simple, but the analysis in M' will give some information about the flow in M.

Finally, it may happen that, outside of a compact subset M_1 of M, the given indeterminacy function $\epsilon(x; u)$ has the property that $\epsilon(x; -v) > \| v \|$, so that the null vector is in the indeterminacy neighborhood at each x of $M - M_1$. This implies that in $M - M_1$ every direction is allowed, so that each component of $M - M_1$ forms a stationary state. Thus the analysis in $M - M_1$ is simple, and the problem is effectively restricted to M_1.

This last case can be interpreted physically as follows. For a particular physical system the accuracy of one's knowledge will usually diminish as one recedes from a given bounded part of the phase space. Thus physical knowledge concerning the cases of extremely large distances or extremely large velocities is very limited. This is reflected in the above condition on the indeterminacy function $\epsilon(x; u)$, for the condition simply states that one is wholly uncertain as to what course the system will take outside of M_1.

6. 2. Applications. The approximation theorem of Section **5** above was developed with particular applications in "non-linear mechanics" in mind, and a number of such problems have been studied from this point of view. It is intended to publish the results in a separate paper. Here some remarks will be made on the methods used and one illustration, a case of the Van der Pol equation, will be given.

In the problems considered, the phase space M was either Euclidean n-space or a product of such a space by a circle. Thus M was non-compact. In all cases either the first or third point of view of Section **6. 1** was applicable, so that one could effectively restrict attention to a compact subset M_1.

Suppose then that M is Euclidean n-space, and that differential equations (1) are given in M. The indeterminacy function $\epsilon_1(x; u)$ may then be given in advance or may be determined at a later stage. The latter procedure is usually more convenient, as will be made clear. One now chooses as M_0 the set of all points $(m_1h, \cdot\cdot\cdot, m_nh)$, where the number $h > 0$ is fixed and the numbers m_i are integers. At each such point x one defines the "first shell" of neighboring points as those points of M_0, other than x, whose coordinates differ from those of x by at most h. Thus, for $n = 2$, there are 8 first shell neighbors: $(m_1 \pm h, m_2h)$, $(m_1, m_2 \pm h)$, $(m_1 \pm h, m_2 \pm h)$. In general, the p-th shell of neighboring points of x is defined as those points of M_0, other than x or those of the first $p - 1$ shells, whose coordinates differ from those of x by at most ph. Thus, for $n = 2$, there are $8p$ p-th shell neighbors.

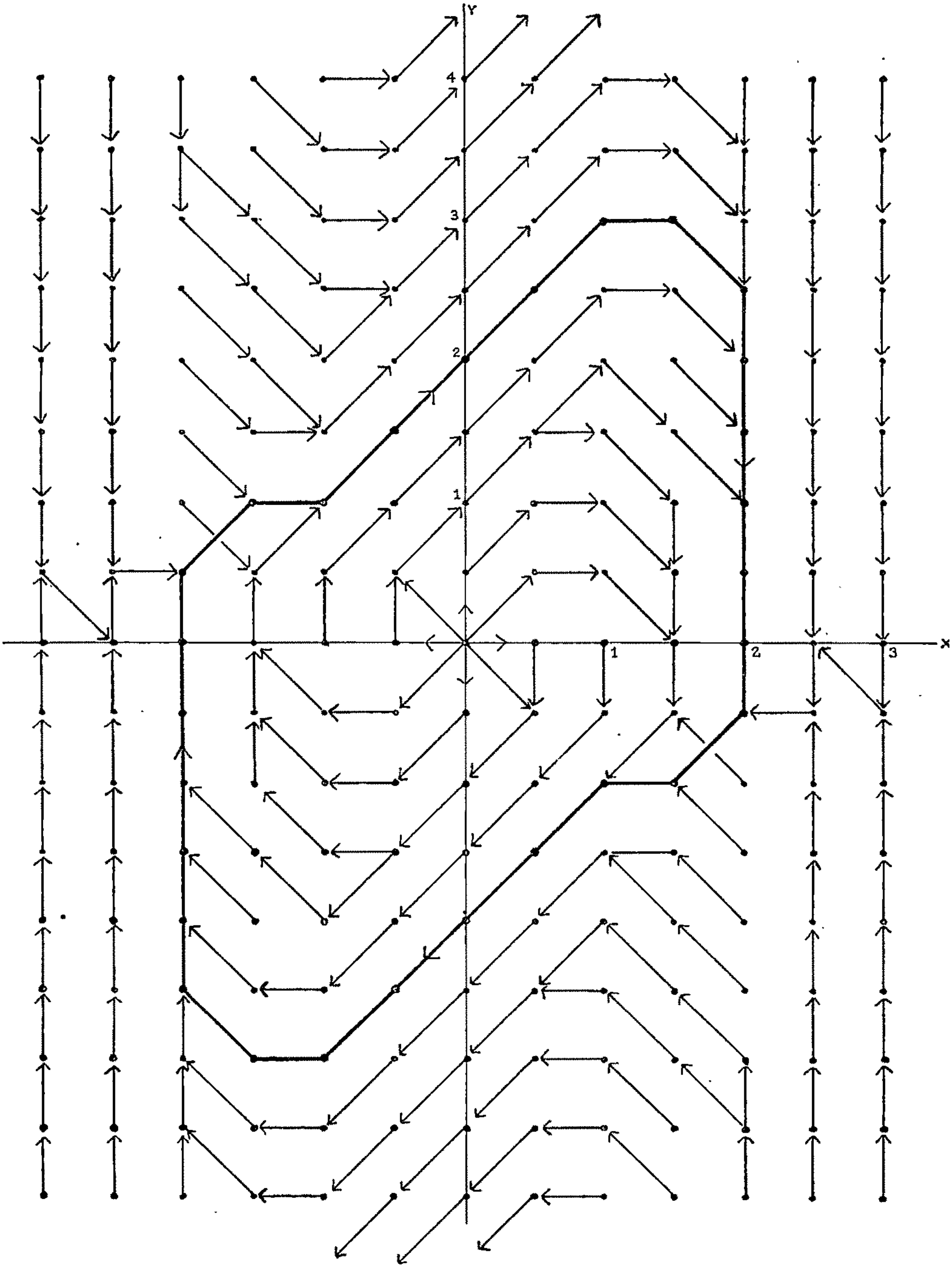
One now chooses a particular value of p, and for each x of M_0 chooses successors only among the points of the first p shells. Fixing h and p is clearly equivalent to fixing the values of δ_1 and δ_2 as in Section **5**. The geodesics are now simply line segments.

If $\epsilon_1(x; u)$ is known, one now chooses as successors of x all points y of the first p shells such that the vector u_{xy}, represented by the line segment xy, satisfies $\| ku_{xy} - v(x) \| < \epsilon_1(x; ku_{xy} - v(x))$ for some $k > 0$. This condition restricts the choice of u_{xy} to a certain cone of directions about that of v.

Usually, however, $\epsilon_1(x; u)$ is not given in advance. One can then effectively choose an ϵ_1 by choosing as successors of x all points y of the first p shells such that the vector u_{xy} makes an angle with v less than a prescribed amount θ. Thus, for $n = 2$ and $p = 1$, one could simply choose all y such that the direction from x to y differs from that of $v(x)$ by less than 23°, so that each x has at least one successor.

Many variations on these procedures are possible. In particular, one can always adjust the $\epsilon_1(x; u)$ to reflect differences in degree of indeterminacy along the directions of the coordinate axes.

Whatever procedure is used, one will obtain a partial order or flow on M_0, and the stable states and complex Φ^0 will be determined as above. The question whether this complex is isomorphic to the complex Φ_1 for the 'corresponding degree of indeterminacy' $\epsilon_1(x; u)$ in M cannot be answered in general. The approximation theorem asserts that this must be the case if ϵ_1 is convex and non-critical and if δ_1, δ_2 and δ_1/δ_2 are sufficiently small (p sufficiently large, h and ph sufficiently small). However, no effective procedure is given for determining when these conditions are met, in particular



Van der Pol equation: $d^2x/dt^2 + (x^2 - 1)dx/dt + x = 0$.
Solutions in the xy-plane, $y = dx/dt$.

for determining when δ_1 and δ_2 are properly chosen. In most particular cases studied thus far, it has been easy to determine when the analysis is fine enough.

As an example, the accompanying figure illustrates the result of applying the method described to the Van der Pol equation: $d^2x/dt^2 + (x^2 - 1)dx/dt + x = 0$. One replaces this by the system $dx/dt = y$, $dy/dt = (1 - x^2)y - x$, and the phase space M is then the xy-plane. The set M_0 is the set of points $(\frac{1}{2}m_1, \frac{1}{2}m_2)$ for integral m_1, m_2. At each point (x, y) the successor (x', y') is chosen in the first shell as the point whose direction differs from the assigned one by less than 23°. Thus, in general at least one successor is obtained. At $(0, 0)$ the vector v reduces to 0, so that the direction is wholly indeterminate.

The figure shows the results of analysis of the rectangle $M_1: |x| \leqq 3$, $|y| \leqq 4$. Outside of this rectangle one can consider the directions as wholly indeterminate, in accordance with the third method of Section 6. 1. If this is done, then one concludes at once that there is one ω-stable state, represented by the heavy simple closed curve. This result agrees with the known properties of the solutions, and the ω-stable state is a reasonably good approximation to the known stable periodic solution. Thus the introduction of indeterminacy in the vector field has not distorted the problem significantly.

QUELQUES PROPRIÉTÉS EXTRÉMALES DU CERCLE ET DE LA SPHÈRE.*

Par Robert Sips.

1. Introduction. Le cercle et la sphère possèdent une série de propriétés isopérimétriques sans relations apparentes les unes avec les autres, mais qui présentent cependant ce caractère commun de ne pouvoir être démontrées par les procédés habituels du calcul des variations.

La plus anciennement connue de ces propriétés se rapporte aux membranes et a été énoncée sans démonstration par Lord Rayleigh [1]. D'après celui-ci, parmi toutes les membranes de même aire et soumises à une même tension uniforme, la membrane circulaire a la fréquence fondamentale la plus basse. Ceci a été démontré pour la première fois par G. Faber [2]. Une autre démonstration a été donnée depuis par E. Krahn [3].

R. Courant [4] a démontré que, parmi toutes les membranes de même longueur périphérique, la membrane circulaire avait la fréquence fondamentale la plus basse.

Une autre propriété extrémale de la sphère a été trouvée par Liapounoff, lequel a démontré que, pour une masse fluide de volume donné, l'énergie potentielle maximum correspondait à la forme sphérique. D'autres démonstrations de cette propriété ont été données par H. Poincaré [5] et par T. Carleman [6].

T. Carleman [7] a démontré que, parmi tous les condensateurs constitués par deux cylindres parallèles indéfinis dont les sections droites ont des aires données, celui constitué par deux cylindres circulaires concentriques a la capacité minimum par unité de longeur.

Ce théorème a été étendu par G. Szegö [8] au cas du condensateur formé par deux surfaces fermées, dont l'une entoure complètement l'autre. La capacité minimum correspond au cas de deux sphères concentriques.

G. Pólya [9], enfin, a prouvé que le module de torsion d'un prisme dont la section droite a une aire donnée est le plus grand possible pour le prisme de section circulaire. Ses résultats ont été généralisés par G. Pólya et A. Weinstein [10].

Dans ce qui suit, nous nous proposons, en généralisant la méthode de

---

* Received November 12, 1949.

G. Szegö, de montrer que toutes les propriétés énumérées plus haut, ainsi que quelques d'autres que nous croyons nouvelles, peuvent être considérées comme des cas particuliers d'une propriété isopérimétrique générale des systèmes de cercles et de sphères concentriques.

2. Inégalité fondamentale. Considérons, dans le plan (x, y) un domaine (D) limité intérieurement par un nombre limité n de courbes fermées extérieures les unes aux autres: $C_{i,1}, C_{i,2}, \cdots, C_{i,n}$, et extérieurement par une courbe fermée C_e qui entoure complètement, sans avoir de points communs avec elles, les courbes C_i. En particulier, C_e peut être située entièrement à l'infini.

Les courbes C_i et C_e seront constituées chacune par un nombre fini d'arcs réguliers et ne possèderont pas de points multiples. L'une ou plusieurs des courbes C_i peuvent d'ailleurs se réduire à un point.

Soit maintenant $z(x, y)$ une solution de l'équation différentielle

$$z_{xx} + z_{yy} = \Delta^2 z = f(z) \tag{1}$$

finie et continue ainsi que ses dérivées partielles des deux premiers ordres dans (D), qui prend la valeur constante z_i sur les courbes C_i et la valeur constante $z_e > z_i$ sur la courbe C_e. Lorsque $z_i \leqq z \leqq z_e$, nous admettrons que $f(z)$ est une fonction analytique positive, bornée et continue de z.

La fonction $z(x, y)$ sera alors une fonction analytique dans tout le domaine (D), sauf peut-être sur les frontières C_i et C_e.

Si nous considérons z comme une troisième coordonnée perpendiculaire au plan des (x, y), l'équation $z = z(x, y)$ représentera une surface dont les courbes de niveau seront les courbes $z =$ const. Nous n'aurons évidemment à considérer que la portion de cette surface dont les points ont leur projection contenue dans (D). Les dérivées partielles z_x et z_y étant partout finies, les courbes de niveau ne pourront avoir de points communs. De plus, $f(z)$ étant par hypothèse toujours positif, la surface ne pourra posséder de maximum à l'intérieur de (D).

Il en résulte qu'une courbe de niveau $z =$ const. sera constituée par un certain nombre de courbes fermées. Lorsque z varie, ce nombre variera chaque fois que z passe par une des valeurs correspondant à un minimum ou à un col de la surface.

Soient maintenant z_1 et z_2 deux constantes comprises entre z_i et z_e. La courbe de niveau complète $z(x, y) = z_1$ se composera d'un certain nombre de courbes fermées. Si z_2 est plus grand et très voisin de z_1, la courbe de niveau $z(x, y) = z_2$ se composera d'un système de courbes fermées peu différentes de celles constituant $z(x, y) = z_1$ et tendant uniformément vers ces dernières lorsque z_2 tend vers z_1.

Dans tout ce qui suit nous supposerons toujours que chacune des courbes fermées partielles du lieu géométrique $z(x, y) = z_2$ entoure complètement la courbe correspondante du lieu $z(x, y) = z_1$ à laquelle elle se réduit lorsque z_2 tend vers z_1.

Ceci exige, comme on peut aisément s'en rendre compte, que la fonction $z(x, y)$ prenne, sur les courbes C_i, une valeur inférieure à celle qu'elle prend sur les courbes de niveau voisines intérieures à (D). Il en résulte que z ne prend jamais, dans (D) de valeur inférieure à z_i et, comme il ne prend pas non plus de valeur supérieure à z_e, il suffira, comme nous l'avons fait plus haut, de considérer les valeurs de z comprises entre z_i et z_e.

Dans le cas particulier où C_e est entièrement située à l'infini, les courbes $z(x, y) = \text{const.} < z_e$ auront tous leurs points à distance finie.

On vérifie immédiatement que lorsque le domaine (D) est constitué simplement par l'aire limitée extérieurement par une courbe fermée C_e, les courbes C_i se réduisent à un ou plusieurs points où z est minimum et la condition restrictive d'enveloppement est toujours vérifiée.

Il résulte de ce qui précède que, lorsque z croît de z_i à z_e, les courbes de niveau $z = z(x, y)$ balayent le domaine (D) d'une manière continue et toujours dans le même sens. Si nous appelons $S(z)$ l'aire totale limitée extérieurement par la courbe $z(x, y) = z$, $S(z)$ sera donc une fonction continue croissante de z.

Appliquons à la région contenue entre $z(x, y) = z_i$ et $z(x, y) = z = \text{const.}$ la première formule de Green

$$(2)\qquad \iint (z_x^2 + z_y^2)\,dS + \iint z\Delta^2 z\,dS = \int_z z\partial z/\partial n\,ds - \int_{z_i} z\partial z/\partial n\,ds,$$

où ds et dS représentent respectivement l'élément d'arc et l'élément d'aire, et où $\partial z/\partial n$ désigne la normale extérieure. Les intégrales curvilignes sont prises le long des courbes de niveau limitant intérieurement et extérieurement le domaine d'intégration.

Si dn est la distance normale entre les courbes voisines z et $z + dz$, nous aurons évidemment

$$(3)\qquad dS/dz = \int_z dn/dz\,ds = \int_z ds/(z_x^2 + z_y^2)^{1/2} = \int_z ds/|\operatorname{grad} z|.$$

D'autre part, en appelant I l'intégrale

$$(4)\qquad I = \iint (z_x^2 + z_y^2)\,dS = \iint |\operatorname{grad} z|^2\,dS;$$

étendue au domaine limité par $z = z_i$ et $z = z$, nous pourrons évidemment écrire

$$(5) \qquad I = \int\int |\operatorname{grad} z|^2 \, dn/dz \, ds \, dz,$$

et par conséquent

$$(6) \qquad dI/dz = \int_z |\operatorname{grad} z| \, ds.$$

En appliquant maintenant au produit $(dS/dz)(dI/dz)$ l'inégalité de Schwarz, il vient

$$(7) \qquad (dS/dz)(dI/dz) \geqq \Big[\int_z ds\Big]^2 = L^2(z),$$

où $L(z)$ est la somme des longueurs des courbes fermées partielles constituant la courbe de niveau $z(x, y) = z$ complète. Si nous tenons compte de l'inégalité isopérimétrique $L^2 \geqq 4\pi S$, (7) a pour conséquence

$$(8) \qquad (dS/dz)(dI/dz) \geqq 4\pi S(z).$$

Mais nous pouvons d'autre part calculer directement la valeur de dI/dz. La relation (2), dérivée par rapport à z, donne en effet

$$(9) \qquad dI/dz = - d\Big(\int_{z_i}^{z} z\Delta^2 z \, dS\Big)/dz + d\Big(\int_z z\partial z/\partial n \, ds\Big)/dz.$$

La première de ces deux intégrales a pour valeur $- z\Delta^2 z \, dS/dz$. Pour calculer la seconde, remarquons que

$$(10) \qquad \int_z z\partial z/\partial n \, ds = z\int_{z_i}^{z} \Delta^2 z \, dS + z\int_{z_i} \partial z/\partial n \, ds,$$

d'où il résulte que

$$(11) \qquad \begin{aligned} dI/dz &= - z\Delta^2 z \, dS/dz + d\Big[z\int_{z_i}^{z} \Delta^2 z \, dS + z\int_{z_i} \partial z/\partial n \, ds\Big]/dz \\ &= \int_{z_i}^{z} \Delta^2 z \, dS + \int_{z_i} \partial z/\partial n \, ds. \end{aligned}$$

Remplaçons maintenant, dans (8), dI/dz par sa valeur (11) ci-dessus et prenons, au lieu de z, S comme variable indépendante. Ceci peut se faire sans ambiguité puisque, comme nous l'avons vu, entre z_i et z_c, S est une fonction toujours croissante de z. Donc dS/dz sera toujours positif, et nous pourrons écrire

$$(12) \qquad \int_{S_i}^{S} \Delta^2 z \, dS \geqq 4\pi S(dz/dS) - \int_{z_i} \partial z/\partial n \, ds,$$

d'où

$$(13) \qquad \int_{S_i}^{S} f(z) \, dS \geqq 4\pi S(dz/dS) - \int_{z_i} \partial z/\partial n \, ds,$$

et enfin

$$dz/dS \leqq (1/4\pi S)\left[\int_{S_i}^{S} f(z)\,dS + \int_{z_i} \partial z/\partial n\, ds\right], \tag{14}$$

ce qui constitue l'inégalité fondamentale que nous désirions établir.

Si l'ensemble des courbes C_i se réduit à une série de points, nous aurons évidemment $S = 0$ pour $z = z_i$. On voit facilement que, dans ce cas, l'inégalité fondamentale devient

$$dz/dS \leqq (1/4\pi S)\cdot \int_{0}^{S} f(z)\,dS, \tag{15}$$

et par conséquent, en particulier, puisque $f(z)$ est une fonction continue bornée

$$\lim_{S=0} dz/dS \leqq (1/4\pi)\cdot f(z_i). \tag{16}$$

Nous avons admis jusqu'ici que la fonction $f(z)$ était une fonction positive de z. Si $f(z)$ est une fonction négative, nous pourrons écrire

$$\Delta^2(-z) = -f(-(-z)), \tag{17}$$

de sorte que l'inégalité (14) s'appliquera à la fonction $-z$. Par conséquent

$$-dz/dS \leqq -(1/4\pi S)\left[\int_{S_i}^{S} f(z)\,dS + \int_{z_i} \partial z/\partial n\, ds\right], \tag{18}$$

d'où

$$dz/dS \geqq (1/4\pi S)\left[\int_{S_i}^{S} f(z)\,dS + \int_{z_i} \partial z/\partial n\, ds\right], \tag{19}$$

c'est-à-dire que le signe de l'inégalité devra être inversé.

Nous allons maintenant rechercher dans quel cas l'inégalité fondamentale devient une égalité. Pour que l'inégalité de Schwarz soit une égalité, il faut que $|\operatorname{grad} z|$ soit constant, c'est-à-dire que les courbes $z = \text{const.}$ soient parallèles. Pour que l'inégalité isopérimétrique soit une égalité, il faut que ces mêmes courbes soient des cercles. La condition nécessaire et suffisante pour que l'inégalité fondamentale devienne une égalité est donc que les courbes de niveau $z = \text{const.}$ constituent un système de cercles concentriques. On aura dans ce cas

$$dz/dS = (1/4\pi S)\left[\int_{S_i}^{S} f(z)\,dS + \int_{z_i} \partial z/\partial n\, ds\right]. \tag{20}$$

Multiplions les deux membres par $4\pi S$, puis dérivons par rapport à S. Il vient

$$4\pi d(S\, dz/dS)/dS = f(z). \tag{21}$$

Prenons comme variable indépendante le rayon r du cercle d'aire S, c'est-à-dire posons $S = \pi r^2$. Ceci donnera à l'équation (21) la forme suivante

$$1/r \cdot d(r\,dz/dr)/dr = f(z), \tag{22}$$

ce qui est, comme on pouvait s'y attendre, l'équation (1) écrite en coordonnées polaires et en supposant que z ne dépend que du rayon r.

Passons maintenant au cas de l'espace à trois dimensions. Nous considérerons un domaine (D) limité intérieurement par n surfaces fermées extérieures les unes aux autres $S_{i,1}, S_{i,2}, \cdots, S_{i,n}$, et extérieurement par une surface fermée S_e entourant complètement les surfaces S_i et n'ayant pas de points communs avec elles.

Les surfaces S_i et S_e seront constituées par un nombre fini de portions régulières. La surface S_e pourra, en particulier, être située entièrement à l'infini.

Nous considérerons ensuite une solution $u(x, y, z)$ de l'équation différentielle

$$u_{xx} + u_{yy} + u_{zz} = \Delta^2 u = f(u) \tag{23}$$

finie et continue dans (D), qui prend la valeur constante u_i sur les surfaces S_i et la valeur constante $u_e > u_i$ sur la surface S_e. Nous admettrons que, lorsque $u_i \leqq u \leqq u_e$, $f(u)$ est une fonction analytique positive, bornée et continue de u.

La fonction $u(x, y, z)$ sera une fonction analytique dans tout le domaine (D), sauf peut-être sur les frontières S_i et S_e. Les surfaces de niveau $u(x, y, z) =$ const. contenues à l'intérieur de (D) seront donc constituées par un certain nombre de surfaces fermées, nombre que dépendra de la valeur particulière de u. Les dérivées partielles u_x, u_y et u_z restant toujours finies, les surfaces de niveau n'auront jamais de points communs. D'autre part, comme, par hypothèse, $f(u)$ est toujours positif, la fonction $u(x, y, z)$ ne pourra avoir de maximum à l'intérieur de (D).

Nous admettrons que, si u_1 et u_2 sont deux valeurs voisines de u, avec $u_2 > u_1$, chacune des surfaces fermées partielles constituant la surface de niveau $u = u_2$ entoure complètement la surface fermée correspondante de $u = u_1$ vers laquelle elle tend lorsque u_2 tend vers u_1.

Ceci exige que la fonction $u(x, y, z)$ prenne, sur les surfaces S_i, une valeur inférieure à celle qu'elle prend sur les surfaces de niveau voisines intérieures à (D). Il en résulte donc que $u(x, y, z)$ sera toujours compris entre u_i et u_e. En particulier, si S_e est située entièrement à l'infini, les surfaces de niveau $u(x, y, z) =$ const. $< u_e$ auront tous leurs points à distance finie.

Lorsque (d) se réduit au volume limité extérieurement par une surface fermée régulière, les surfaces S_i sont constituées par le point ou les points où $u(x, y, z)$ prend sa valeur minimum. On vérifie alors immédiatement que la condition d'enveloppement est toujours satisfaite.

Il résulte de ce qui précède que, lorsque u croît de u_i à u_e, les surfaces $u(x, y, z) = u$ balayent le domaine (D) d'une manière continue et toujours dans le même sens. Si nous appelons $V(u)$ le volume total limité extérieurement par la surface de niveau $u(x, y, z) = u$, V sera donc une fonction continue croissante de u.

L'inégalité fondamentale s'établira à peu près exactement comme dans le cas de deux dimensions. La première formule de Green, appliquée à la région comprise entre $u(x, y, z) = u_i$ et $u(x, y, z) = u = \text{const.}$, donnera tout d'abord

$$(24)\qquad \int\!\!\int\!\!\int (u_x^2 + u_y^2 + u_z^2)\,dV + \int\!\!\int\!\!\int u\Delta^2 u\,dV = \int\!\!\int_{u_i} u\partial u/\partial n\,dS - \int\!\!\int_{u} u\partial u/\partial n\,dS,$$

où dV et dS représentent respectivement l'élément de volume et l'élément de surface, et où $\partial u/\partial n$ désigne la dérivée normale extérieure. Les intégrales doubles sont prises le long des surfaces de niveau limitant intérieurement et extérieurement le domaine d'intégration.

En appelant I l'intégrale

$$(25)\qquad I = \int\!\!\int\!\!\int (u_x^2 + u_y^2 + u_z^2)\,dV,$$

étendue à la même région, on démontrera ensuite que, en vertu de l'inégalité de Schwarz

$$(26)\qquad (dV/du)\,(dI/du) \geqq S^2(u),$$

où $S(u)$ représente la somme des aires des surfaces fermées constituant la surface de niveau $u(x, y, z) = u$. Si l'on applique maintenant l'inégalité isopérimétrique $S(u) \geqq [36\pi V^2(u)]^{1/3}$, il vient

$$(27)\qquad (dV/du)\,(dI/du) \geqq [36\pi V^2]^{2/3}.$$

On établira d'autre part que

$$(28)\qquad dI/du = \int_{V_i}^{V} \Delta^2 u\,dV + \int\!\!\int_{S_i} \partial u/\partial n\,dS,$$

et par conséquent, enfin, que

$$(29)\qquad [36\pi V^2]^{2/3} du/dV \leqq \int_{V_i}^{V} f(u)\,dv + \int\!\!\int_{S_i} \partial u/\partial n\,dS,$$

ce qui constitue l'inégalité fondamentale que nous voulions démontrer, dans le cas de trois dimensions.

On voit immédiatement que des inégalités analogues pourraient être obtenues de la même manière dans le cas d'un nombre supérieur de dimensions.

On démontrera comme plus haut que, pour que l'inégalité fondamentale devienne une égalité, il est nécessaire et suffisant que les surfaces $u(x, y, z) = \text{const.}$ constituent un système de sphères concentriques. Nous aurons donc, dans ce dernier cas

$$(30) \qquad [36\pi V^2]^{1/3} du/dV = \int_{V_i}^{V} f(u)\,dV + \int\int_{S_i} \partial u/\partial n\, dS.$$

Posons $V = 4/3\pi r^3$, puis dérivons les deux membres par rapport à r. Il vient

$$(31) \qquad 1/r^2 \cdot d(r^2 du/dr)/dr = f(u),$$

ce qui n'est autre que l'équation (23) écrite en coordonnées polaires et en admettant que u est seulement fonction de r.

Si la surface S_i se réduit à une série de points, on aura évidemment $V = 0$ pour $u = u_i$. L'inégalité (29) prend alors la forme

$$(32) \qquad [36\pi V^2]^{1/3} du/dV \leqq \int_0^V f(u)\,dV,$$

d'où l'on déduit que

$$(33) \qquad \lim_{V=0} V^{1/3} du/dV \leqq (36\pi)^{-2/3} f(u_i).$$

On démontrera enfin facilement que si $f(u)$, au lieu d'être une fonction positive, est une fonction toujours négative, il faudra changer le sens de l'inégalité fondamentale qui deviendra ainsi

$$(34) \qquad (36\pi V^2)^{1/3} du/dV \geqq \int_{V_i}^{V} f(u)\,dV + \int\int_{S_i} \partial u/\partial n\, dS.$$

3. Interprétation géométrique de l'inégalité fondamentale. L'inégalité fondamentale est susceptible d'une interprétation géométrique intéressante.

Admettons que, dans le cas de deux dimensions, $f(z)$ est une fonction croissante de z et que, pour $z = z_i$, l'aire $S(z_i)$ des courbes C_i et l'intégrale $\int_{z_i} (\partial z/\partial n)\,ds$ étendue à ces mêmes courbes aient une valeur fixe, indépendante du système particulier de courbes C_i considéré.

La courbe $z = z(S)$ passera alors toujours par le point fixe $z = z_i$, $S = S_i$, et aura son coefficient angulaire, en tout point, inférieur à la quantité

$$(4\pi S)^{-1}\Big[\int_{S_i}^{S} f(z)\,dS + \int_{z_i} \partial z/\partial n\, ds\Big].$$

Appelons $Z(S)$ la solution de l'équation (20) qui passe également par le point z_i, S_i. Nous aurons alors

$$dZ/dS - dz/dS \geqq \int_{S_i}^{S} (f(Z) - f(z))\,dS. \tag{35}$$

Or, aux environs immédiats du point z_i, S_i, on a évidemment $d(Z-z)/dS \geqq 0$, d'où $Z \geqq z$ et $f(Z) \geqq f(z)$. En raisonnant de proche en proche, on voit qu'on aura toujours $d(Z-z)/dS \geqq 0$ et $Z \geqq z$. Il en résulte que la courbe $Z(S)$ sera toujours située au dessus de la courbe $z(S)$ et que l'écart entre ces deux courbes augmentera en même temps que S.

La courbe $Z(S)$ fournit donc, pour chaque valeur de z, une limite inférieure pour l'aire limitée par la courbe de niveau $z(x,y) = z$.

Dans le cas de trois dimensions, on peut établir une propriété analogue.

Nous admettrons encore une fois que $f(u)$ est une fonction toujours positive et croissante de u, et que le volume V_i limité par les surfaces S_i ainsi que l'intégrale $\int\int_{S_i} \partial u/\partial n\, dS$ ont une valeur fixe, indépendante du système particulier de surfaces S_i. Pour $u = u_i$, on aura donc toujours $V = V_i$. Appelons $U(V)$ la solution de l'équation (30) passant également par le point $u = u_i$, $V = V_i$. On aura évidemment

$$dU/dV - du/dV \geqq (36\pi V^2)^{-2/3} \int_{V_i}^{V} (f(U) - f(u))\,dV, \tag{36}$$

et par conséquent la courbe $U(V)$ se trouvera toujours au dessus de la courbe $u(V)$. La courbe $U(V)$ permettra donc, pour toutes les valeurs de u, d'obtenir une limite inférieure du volume V limité par la surface $u(x,y,z) = u$.

Nous allons maintenant appliquer les inégalités fondamentales (14) et (29) à une série de cas particuliers correspondant à différentes formes de la fonction f.

4. Condensateur électrostatique de capacité minimum. Considérons un condensateur constitué par deux surfaces conductrices fermées, telles que la surface extérieure S_e entoure complètement, sans la toucher, la surface intérieure S_i. Ces deux surfaces sont portées respectivement aux potentiels u_e et u_i, avec $u_e > u_i$.

Dans le domaine compris entre les deux surfaces, le potentiel $u(x,y,z)$ vérifie l'équation de Laplace $\Delta^2 u = 0$, et l'intégrale $\int \partial u/\partial n\, dS$ prise sur l'une quelconque des surfaces équipotentielles comprises entre S_i et S_e a une valeur constante et égale à $4\pi Q$, où Q est la charge de l'armature intérieure.

On voit immédiatement que les conditions d'application de l'inégalité

(28) sont remplies, et celle-ci devient, dans le cas actuel, en posant $V = 4\pi r^3/3$, l'inégalité $r^2 du/dr \leqq Q$. Divisons les deux membres de cette inégalité par r^2 et intégrons ensuite entre r_i et r, avec bien entendu, $V_i = 4\pi r_i^3/3$. En tenant compte de ce que, pour $u = u_i$, on a $r = r_i$, il vient $u \leqq u_i + Q(1/r_i - 1/r)$, et en particulier, sur l'armature extérieure, où $r = r_e$, $u_e \leqq u_i + Q(1/r_i - 1/r_e)$. En appelant C la capacité du condensateur, ceci peut encore s'écrire

$$C = Q/(u_e - u_i) \geqq 1/(1/r_i - 1/r_e) = (3/4\pi)^{1/3}/(V_i^{-1/3} - V_e^{-1/3}),$$

c'est-à-dire que:

Parmi tous les condensateurs électrostatiques tels que les armatures extérieures et intérieures entourent des volumes donnés, celui constitué par deux sphères concentriques a la plus petite capacité.

On peut évidemment faire un raisonnement analogue dans le cas d'un condensateur électrostatique constitué par deux armatures cylindriques parallèles telles que l'une entoure complètement l'autre et dont les sections droites ont des aires données. La capacité par unité de longueur minimum correspond alors au condensateur constitué par deux cylindres circulaires coaxiaux.

Dans le cas du condensateur de capacité minimum, notre méthode est évidemment identique à celle utilisée par G. Szegö dans le travail déja mentionné plus haut.

5. Application à la théorie des membranes. La déformation normale d'une membrane plane uniforme et uniformément tendue soumise à une pression constante est donnée par l'équation $\Delta^2 z = -p/T$, où T est la tension et p la pression par unité de surface.

Supposons la membrane limitée par un contour fermé sans points multiples. Sur ce contour on devra avoir $z = 0$. Les courbes de niveau $z =$ const. seront également des courbes fermées qui entourent le point ou les points où z est maximum. Prenons ces points pour origine des S et posons $z = z_M - z_1$, où z_M représente la valeur maximum de z.

La quantité z_1 vérifiera l'équation $\Delta^2 z_1 = p/T$, et on voit sans peine que l'inégalité fondamentale pourra être appliquée au système des courbes $z_1 =$ const. Dans le cas présent, (14) devient simplement $dz_1/dS \leqq p/4\pi T$, d'où, en intégrant entre $z_1 = 0$ et $z_1 = z_M$, $z_M \leqq pS_e/4\pi T$, c'est-à-dire que:

Parmi toutes les membranes de même aire, soumises à la même tension, et supportant la même pression uniforme, la membrane circulaire présente la plus grande déformation normale maximum.

6. Application à la théorie de la torsion des prismes. L'étude de la torsion d'un prisme droit de section quelconque se ramène, comme on le sait, à la détermination de la solution de l'équation $\Delta^2 z = -2$, qui s'annule sur le contour de la section droite. Nous supposerons celle-ci limitée par une courbe fermée sans points multiples. Nous pourrons appliquer les résultats du paragraphe précédent et, en posant encore une fois $z = z_M - z_1$, écrire

$$dz_1/dS \leqq 1/2\pi. \tag{37}$$

L'angle de rotation du prisme, par unité de longueur, a pour valeur $\vartheta = M/C$, où M est le couple de torsion et C le module de torsion donné par $C = 2G \int z\, dS$, G étant le module de glissement et l'intégrale étant étendue à l'aire de la section droite. On vérifie aisément qu'on a également $C = 2G \int S\, dz_1$, et par conséquent, en vertu de (37)

$$C = 2G \int S\, dz_1 \leqq 2G \int_0^{S_e} S\, dS/2\pi = GS_e^2/2\pi, \tag{38}$$

S_e étant ici l'aire de la section droite. Ceci montre que:

Parmi tous les prismes de même nature et dont la section droite a la même aire, le prisme de section circulaire possède le module de torsion le plus élevé.

7. Application à l'hydrodynamique des liquides visqueux. On démontre aisément, par un calcul identique à celui qui précède, que le débit d'un liquide visqueux s'écoulant en régime laminaire dans un tube cylindrique de section S_0 est soumis à l'inégalité suivante: $Q \leqq (P\rho/8\pi\mu L) S_0^2$, où Q est le débit, P la différence de pression aux extrémités du tube, ρ la densité, μ le coefficient de viscosité. L'égalité n'a lieu que pour une section circulaire.

Donc, si l'on fait écouler, en régime laminaire et sous l'influence d'une même différence de pression par unité de longueur, un même liquide visqueux dans des tubes cylindriques de même longueur, on obtiendra le débit maximum avec un tube de section circulaire.

8. Problème de Rayleigh. La fonction caractéristique fondamentale de l'équation des membranes vibrantes

$$\Delta^2 z + k^2 z = 0, \tag{39}$$

relativement à une courbe fermée (C) sans points multiples est la solution $z(x, y)$ de cette équation, toujours positive, nulle sur le contour. La valeur correspondante de k est, à un facteur constant près, la fréquence fondamentale.

D'après les propriétés de l'équation (39), la fonction $z(x,y)$ ne peut pas posséder de minimum à l'intérieur de la courbe (C), mais elle peut posséder plusieurs maxima distincts. Appelons z_M la plus grande valeur de $z(x,y)$ dans (C), cette valeur pouvant être atteinte en plusieurs points distincts. Les courbes de niveau $z = \text{const.}$, lorsque z varie entre 0 et z_M, forment un système de courbes fermées analytiques sans points communs s'enveloppant mutuellement. Posons $z_1 = z_M - z$. La fonction $z_1(x,y)$ vérifiera l'équation $\Delta^2 z_1 = k^2(z_M - z_1)$, dont le second membre sera toujours positif. Nous pourrons donc appliquer l'inégalité fondamentale (15) qui deviendra dans le cas actuel

$$dz_1/dS \leqq (k^2/4\pi S)\int_0^S (z_M - z_1)\,dS. \tag{40}$$

Si S_0 représente l'aire de la membrane, lorsque S variera entre S et S_0, z_1 sera une fonction croissante de S et dz_1/dS sera positif. En multipliant les deux membres de (40) par $S\,dz_1/dS$ et en intégrant entre 0 et S_0, on aura donc

$$\int_0^{S_0} S(dz_1/dS)^2 dS \leqq (k^2/4\pi)\int_0^{S_0} dz_1/dS\left[\int_0^S (z_M - z_1)\,dS\right] dS.$$

Le second membre peut être intégré par parties et devient

$$\int_0^{S_0} dz_1/dS\left[\int_0^S (z_M - z_1)\,dS\right] dS = \left[z_1 \int_0^S (z_M - z_1)\,dS\right]_0^{S_0} - \int_0^{S_0} z_1(z_M - z_1)dS$$

$= \int_0^{S_0} (z_M - z_1)^2 dS$ de sorte que, en revenant à la variable z, on obtient

$$\int_0^{S_0} S(dz/dS)^2 dS \leqq (k^2/4\pi)\int_0^{S_0} z^2 dS.$$

Posons maintenant $S = \pi r^2$. Il vient

$$\int_0^{r_0} (dz/dr)^2 r\,dr \leqq k^2 \int_0^{r_0} z^2 r\,dr, \tag{41}$$

avec $S_0 = \pi r_0^2$. Or $z(r)$ est une fonction continue et dérivable de r, pour $0 \leqq r \leqq r_0$, qui s'annule pour $r = r_0$. On pourra donc la développer en série convergente et dérivable de fonctions de Bessel $J_0(\lambda_n r)$, où les constantes λ_n sont définies par la condition

$$J_0(\lambda_n r_0) = 0. \tag{42}$$

On aura donc $z = \Sigma_n A_n J_0(\lambda_n r)$, où les A_n sont des constantes. En introduisant cette valeur de z dans l'inégalité (41), celle-ci devient

$$\frac{\int_0^{r_0}(dz/dr)^2 r\,dr}{\int_0^{r_0} z^2 r\,dr} = \frac{\Sigma A_n^2\lambda_n^2 \int_0^{r_0} rJ_0^2(\lambda_n r)\,dr}{\Sigma A_n^2 \int_0^{r_0} rJ_0^2(\lambda_n r)\,dr} \leqq k^2.$$

Les quantités λ_n forment une suite de nombres croissant indéfiniment. En appelant $\lambda_0 r_0$ la plus petite des racines de (42), on aura donc $k^2 \geqq \lambda_0^2 = (2{,}405\cdots)^2\pi/S_0$ $(\lambda_0 r_0 = 2{,}405\cdots)$ c'est-à-dire que la fréquence fondamentale d'une membrane non circulaire sera toujours supérieure à la fréquence fondamentale de la membrane circulaire de même aire.

On peut démontrer de la même manière la propriété correspondante dans le cas de trois dimensions.

9. Problème de Liapounoff. Considérons une masse fluide incompressible, immobile, de volume V_0, en équilibre sous l'influence de son attraction newtonienne propre. On sait que, dans ces conditions, la surface extérieure sera une surface équipotentielle.

Nous devrons considérer séparément le potentiel newtonien u_i à l'intérieur de la masse et le potentiel u_e à l'extérieur de celle-ci. Ces deux fonctions vérifient les équations suivantes $\Delta^2 u_i = -4\pi k\rho$, $\Delta^2 u_e = 0$, où ρ est la densité (constante) du fluide et k la constante de gravitation. A la surface même du liquide, on devra évidemment avoir $u_i = u_e = u_0$.

On voit immédiatement que, à l'extérieur de la masse fluide, l'inégalité fondamentale (33) pourra être appliquée aux surfaces équipotentielles $u_e = \text{const}$. Par conséquent

$$(36\pi V^2)^{2/3} du_e/dV \geqq \int \partial u_e/\partial n\, dS = -4\pi k\rho V_0.$$

En posant $V = 4\pi r^3/3$, ceci peut encore s'écrire $du_e/dr + V_0 k\rho/r^2 \geqq 0$. Intégrons le premier membre de cette inégalité entre r et l'infini. Pour V très grand, les surfaces équipotentielles diffèrent très peu de sphères, de sorte que, dans ce cas $u_e = V_0 k\rho/r +$ termes en $r^{-2}, r^{-3}, \cdots$ et par conséquent $u_e - V_0 k\rho/r$ s'annule pour $r = \infty$. Il reste alors $-(u_e - V_0 k\rho/r) \geqq 0$, ou bien

$$u_e \leqq V_0 k\rho/r. \tag{43}$$

Considérons ensuite le potentiel u_i à l'intérieur de la masse fluide. Celui-ci prendra évidemment une valeur maximum en un ou plusieurs points, mais ne possèdera certainement pas de minimum. Les surfaces équipotentielles, coincidant avec les surfaces d'égale pression, s'enveloppent mutuellement et nous pourrons appliquer l'inégalité fondamentale (33) en prenant pour origine $V=0$ le point ou les points où u_i est maximum. Nous obtenons ainsi $(36\pi V^2)^{2/3} du_i/dV \geqq -4\pi k\rho V$, ce qui peut encore s'écrire $du_i/dr + 4/3\cdot\pi k\rho r \geqq 0$, en posant comme plus haut, $V=4\pi r^3/3$. Intégrons cette inégalité entre r et r_0. Il vient

$$u_0 - u_i + 2/3\cdot\pi k\rho r_0^2 - 2/3\cdot\pi k\rho r^2 \geqq 0\,,$$

d'où

$$u_i \leqq u_0 + 2/3\cdot\pi k\rho r_0^2 - 2/3\cdot\pi k\rho r^2,$$

et, en tenant compte de (43), $u_i \leqq 2\pi k\rho r_0^2 - 2/3\cdot\pi k\rho r^2$. Or l'énergie potentielle de la masse a pour valeur

$$E = \tfrac{1}{2}\rho\int_0^{V_0} u_i dV = 2\pi\rho\int_0^{r_0} u_i r^2 dr,$$

et par conséquent, enfin

$$E \leqq 16/15\cdot\pi^2 k\rho^2 r_0^5 = 16/15\cdot\pi^2 k\rho^2 (3V_0/4\pi)^{5/3}.$$

Si donc il existe une forme d'équilibre non sphérique, son énergie potentielle gravitationnelle sera inférieure à celle de la forme sphérique d'équilibre de même volume.

Il en résulte donc que la sphère constitue la forme d'équilibre stable d'une masse fluide incompressible immobile.

Considérons maintenant une masse fluide constituée par des volumes V_1 et V_2 de deux liquides immiscibles différents, de densités constantes ρ_1 et ρ_2, avec $\rho_1 > \rho_2$. A l'état d'équilibre, on sait que les surfaces équipotentielles coincident avec les surfaces d'égale pression et les surfaces d'égale densité. D'autre part, pour que l'équilibre soit stable, il faut évidemment que la pression augmente au fur et à mesure que l'on s'éloigne de la surface extérieure. Nous ne considérerons que les formes d'équilibre où le liquide le moins dense entoure complètement le liquide le plus dense. La surface de séparation, de même que la surface extérieure, seront des surfaces équipotentielles.

Le potentiel newtonien de la masse totale sera une fonction continue et à dérivées partielles continues des coordonnées. Ce sera donc aussi une fonction continue et à dérivée continue du volume V limité par la surface équipotentielle $u(x,y,z) = \text{const.}$ On aura de plus:

Pour $V > V_1 + V_2$, $\Delta^2 u = 0$; pour $V_1 + V_2 > V > V_1$, $\Delta^2 u = -4\pi k\rho_2$; et pour $V_1 > V$, $\Delta^2 u = -4\pi k\rho_1$.

En introduisant ces valeurs dans l'inégalité fondamentale et en procédant comme ci-dessus, on démontrera aisément qu'une configuration d'équilibre éventuelle, différente de celle constituée par une sphère de densité ρ_1 et de volume V_1, entourée d'une couche sphérique concentrique de densité ρ_2 et de volume V_2 possèdera une énergie potentielle inférieure à celle de la configuration sphérique concentrique. Celle-ci constituera donc certainement la forme d'équilibre stable de la masse.

Nous pourrons étudier de la même manière la forme d'équilibre d'une masse hétérogène constituée par des volumes $V_1, V_2, \cdots, V_n$ de n fluides incompressibles immiscibles de densités $\rho_1, \rho_2, \cdots, \rho_n$. Si nous supposons que $\rho_1 > \rho_2 \cdots > \rho_n$, nous pourrons, en utilisant toujours le même procédé, démontrer que la configuration d'équilibre stable sera constituée par une sphère de volume V_1 et de densité ρ_1, entourée d'une série de couches sphériques concentriques de densités décroissantes ayant respectivement les volumes $V_2, \cdots, V_n$ et les densités $\rho_2, \cdots, \rho_n$.

Ce dernier résultat peut aisément être généralisé au cas d'une variation continue de la densité.

Les résultats qui précèdent ont été obtenus d'une manière entièrement différente par L. Lichtenstein [11].

10. Inégalités relatives à la fonction de Green. Considérons, dans le plan, un contour fermé sans points multiples (C) et soit $P(a, b)$ un point intérieur quelconque fixe. La fonction de Green correspondante $G(a, b, x, y) = G(P, Q)$ s'annulera sur le contour et aura la forme suivante: $G(a, b, x, y) = \log \cdot r_{PQ} + H(a, b, x, y)$, où r_{PQ} est la distance entre les points $P(a, b)$ et $Q(x, y)$, et H est une fonction harmonique de x et de y. La fonction $G(P, Q)$ vérifie évidemment l'équation $\Delta^2 G = 0$, lorsque $Q \neq P$. Les courbes $G(P, Q) = \text{const.} = G$ sont des courbes analytiques, (sauf peut-être pour $G = 0$) fermées s'enveloppant mutuellement. Pour $G = -\infty$, la courbe se réduit au point P, tandis que pour $G = 0$, elle coincide avec (C). L'intégrale curviligne $\int \partial G/\partial n \, ds$ prise le long d'une des courbes fermées $G = \text{const.} > -\infty$ a une valeur constante et égale à 2π. It est donc possible d'appliquer l'inégalité fondamentale au système des courbes $G = \text{const.}$ et on obtient ainsi $dG/dS \leqq 1/2S$. Comme G croît en même temps que S, la dérivée sera positive et nous pourrons intégrer les deux membres de l'inégalité entre S et S_0, S_0 étant l'aire limitée par la courbe (C). Puisque, pour $G = 0$, on a $S = S_0$, on obtient ainsi $-G \leqq \frac{1}{2} \log S_0/S$, ce qu'on peut encore écrire $S \leqq S_0 \exp 2G$.

L'aire limitée par la courbe $G(x, y) = \text{const.}$ sera donc inférieure à $S_0 \exp. 2G$, sauf évidemment si (C) est un cercle dont P est le centre.

Dans le cas de trois dimensions, nous aurons à considérer le domaine limité par une surface régulière fermée (S). La fonction de Green aura la forme $G(P, Q) = -1/r_{PQ} + H(P, Q)$, où r_{PQ} est la distance du point fixe

$P(a, b)$ au point variable $Q(x, y)$ et H une fonction harmonique, et s'annulera sur la surface (S). L'intégrale de surface $\int\int(\partial G/\partial n)dS$ prise sur l'une quelconque des surfaces fermées analytiques $G = \text{const.} > -\infty$, aura la valeur constante 4π.

Il sera de nouveau possible d'appliquer l'inégalité fondamentale qui donnera maintenant $dG/dV \leqq 4\pi(36\pi V^2)^{-2/3}$. En intégrant entre V et V_0, où V_0 est le volume limité par la surface (S), il viendra $G \geqq (4\pi/3)^{1/3}$ $(V_0^{-1/3} - V^{-1/3})$, puisque $V = V_0$ pour $G = 0$. L'inégalité ne devient une égalité que lorsque (S) est une sphère de centre P (sauf bien entendu pour les valeurs exceptionnelles $G = 0$ et $G = -\infty$).

D'autres inégalités relatives à la fonction de Green dans le plan peuvent être obtenues par simple changement de forme des résultats donnés plus haut.

On vérifie facilement que la solution de l'équation $\Delta^2 z = -1$ régulière à l'intérieur du contour (C) et s'annulant sur celui-ci est donnée par

$$z(x, y) = 1/2\pi \int\int G(x, y, \xi, \eta)\, d\xi\, d\eta,$$

de sorte que l'inégalité $z_M \leqq pS_e/4\pi T$, trouvée à la fin du § **5**, peut s'écrire

$$\int\int G(x, y, \xi, \eta)\, d\xi\, d\eta \leqq \tfrac{1}{2} S_0. \tag{44}$$

De même, l'inégalité (38) peut se mettre sous la forme suivante

$$\int\int\int\int G(x, y, \xi, \eta)\, dx dy d\xi d\eta \leqq \tfrac{1}{4} S_0^2.$$

Dans ses travaux relatifs à la résolution du problème de Dirichlet pour les équations de la forme $\Delta^2 u = c(x, y)u$ et $\Delta^2 u = F(x, y, u)$, E. Picard [12] a rencontré l'inégalité suivante

$$1/2\pi \int\int G(x, y, \xi, \eta)\, d\xi d\eta \leqq \lambda,$$

et a démontré que λ tendait vers zéro en même temps que l'aire limitée par le contour. Notre formule (44) donne pour λ la valeur exacte $S_0/4\pi$. Il n'y a aucune difficulté à obtenir des résultats analogues dans le cas de trois dimensions. On trouve ainsi

$$\int\int\int G(x, y, z, \xi, \eta, \zeta)\, d\xi d\eta d\zeta \leqq 1/6 \cdot (3V_0/4\pi)^{2/3},$$

$$\int\int\int\int\int\int G(x, y, z, \xi, \eta, \zeta)\, dx dy dz d\xi d\eta d\zeta \leqq 4\pi/45 \cdot (3V_0/4\pi)^{5/3}.$$

Encore une fois, ces inégalités ne deviennent des égalités que lorsque la frontière est un cercle ou une sphère.

Note additionelle.[1] Un examinateur nous a indiqué une démonstration de l'inégalité (12) beaucoup plus simple que celle donnée plus haut. En utilisant la même notation que dans le paragraphe **2**, on voit aisément que

---

[1] Received April 28, 1950.

$$\int_z \partial z/\partial n\, ds \times \int_z ds/(\partial z/\partial n) > [\int ds]^2 = L^2(z) > 4\pi S(z).$$

Mais $\int_z ds/(\partial z/\partial n) = dS/dz$, de sorte que $\int_z \partial z/\partial n\, ds \times dS/dz \geqq 4\pi S(z)$, ou, en prenant S comme variable indépendante, $\int_z \partial z/\partial n\, ds \geqq 4\pi S . dz/dS$. Si maintenant nous faisons usage de la formule de Green

$$\int\int \Delta^2 z\, dS = \int_z (\partial z/\partial n)\, ds - \int_{z_1} \partial z/\partial n\, ds,$$

nous obtenons finalement l'inégalité (12),

$$\int\int \Delta^2 z\, dS + \int_{z_1} \partial z/\partial n\, ds \geqq 4\pi S\, dz/dS.$$

Brussels, Belgium.

BIBLIOGRAPHIE.

[1] J. W. S. Rayleigh, *Theory of sound*, Second edition, vol. 1, London, 1894, p. 341.

[2] G. Faber, "Beweis, dass unter allen homogenen Membranen von gleicher Fläche und gleicher Spannung die kreisförmige den tiefsten Grundton gibt," *Sitzungsberichte der mathematischen-physikalischen Klasse der Bayerischen Akademie der Wissenschaften*, 1923, pp. 169-172.

[3] E. Krahn, "Ueber eine von Rayleigh formulierte Minimaleigenschaft des Kreises," *Mathematische Annalen*, vol. 94 (1925), pp. 97-100.

[4] R. Courant, "Beweis des Satzes, dass von allen homogenen Membranen gegebenen Umfanges und gegebener Spannung die kreisförmige den tiefsten Grundton besitzt," *Mathematische Zeitschrift*, vol. 1 (1918), pp. 321-328.

[5] H. Poincaré, *Figures d'équilibre d'une masse fluide*, Paris, 1902, p. 15.

[6] T. Carleman, "Ueber eine isoperimetrische Aufgabe und ihre physikalischen Anwendungen," *Mathematische Zeitschrift*, vol. 3 (1919), pp. 1-7.

[7] ———, "Ueber ein Minimalproblem der mathematischen Physik," *ibid.*, vol. 1 (1918), pp. 208-212.

[8] G. Szegö, "Ueber einige Extremalaufgaben der Potentialtheorie," *ibid.*, vol. 31 (1930), pp. 583-593.

[9] G. Pólya, "Torsional rigidity, principal frequency, electrostatic capacity, and symmetrization," *Quarterly of Applied Mathematics*, vol. 6 (1948), pp. 267-277.

[10] ——— and A. Weinstein, "Approximations and bounds for the electrostatic capacity and for similar physical quantities," *Bulletin of the American Mathematical Society*, vol. 55 (1949), p. 294.

[11] L. Lichtenstein, *Gleichgewichtsfiguren rotierender Flüssigkeiten*, Berlin, 1933, p. 26.

[12] É. Picard, *Leçons sur quelques problèmes aux limites de la théorie des équations différentielles*, Paris, 1940, pp. 131-132.

ON DENUMERABLY INDEPENDENT FAMILIES OF BOREL FIELDS.*

By S. SHERMAN.

In a posthumous paper [B] Banach showed that if one has a denumerably independent family of Borel fields (each with a measure) of subsets of a set Ω, then there exists a common extension of the measures to the Borel field generated by the Borel fields of the family such that the fields are stochastically independent. Since his proof is quite long we give a shorter proof which has the advantage that it displays the relation between a denumerably independent family of Borel fields and an isomorphic family of Borel fields in a product space. As a consequence of this the existence (and uniqueness) of Banach's extension measure is equivalent to the existence (and uniqueness) of an independent product measure, a solved problem.[1,2]

Let Ω be a set and let $\mathfrak{B}^\gamma = \{B^\gamma\}$ be a Borel field of subsets of Ω for each $\gamma \varepsilon \Gamma$. Let $\mathfrak{A} = \{\mathfrak{B}^\gamma \mid \gamma \varepsilon \Gamma\}$. We say that $E[H]$ is a $\mathfrak{A}$-rectangle [$\mathfrak{A}$-block] if $E[H] = \bigcap_\gamma B^\gamma$, where all except a finite [countable] collection of the B^γ are equal to Ω. In the sequel H (with or without subscripts) denotes an $\mathfrak{A}$-block. Let $\mathfrak{E}$ [$\mathfrak{B}$] be the [Borel] field generated by the rectangles. We note that i) if $A \varepsilon \mathfrak{E}$ then A is a finite disjoint union of rectangles, ii) if E is an $\mathfrak{A}$-rectangle then E is an $\mathfrak{A}$-block, and iii) $\mathfrak{B}$ is the smallest Borel field containing each $\mathfrak{B}^\gamma \varepsilon \mathfrak{A}$.

LEMMA. *If* $A \varepsilon \mathfrak{E}$, *then*

$*)$ $\quad p \varepsilon A \Longrightarrow$ for some $\underline{H} \subset A$, $p \varepsilon \underline{H}$,

$**)$ $\quad p \varepsilon A' \Longrightarrow$ for some $\bar{H} \subset A'$, $p \varepsilon \bar{H}$.

(Note: A' means the complement of A.)

Proof. Trivial since both A and A' are disjoint unions of $\mathfrak{A}$-rectangles.

---

* Received October 24, 1949; revised January 20, 1950.

[1] See [*K*] where a recent proof as well as references to the literature are given. See also [*SJ*].

[2] According to Mr. Henry Helson another (different) proof by Saks of Banach's theorem is to appear in a Polish Journal.

THEOREM. *If $A \varepsilon \mathfrak{B}$, then * and **.*

Proof. If A satisfies * and **, then A' satisfies * and **. Let $A = \bigcap_{\alpha=1}^{\infty} B_\alpha$, where for each α,

$$p \varepsilon B_\alpha \Longrightarrow \text{ for some } \underline{H}_\alpha \subset B_\alpha,\ p \varepsilon \underline{H}_\alpha,$$

$$p \varepsilon B'_\alpha \Longrightarrow \text{ for some } \bar{H}_\alpha \subset B'_\alpha,\ p \varepsilon \bar{H}_\alpha.$$

If $q \varepsilon A$, then $q \varepsilon \bigcap_{\alpha=1}^{\infty} \underline{H}_\alpha$, which is an $\mathfrak{A}$-block. If $q \varepsilon A'$, then for some $\bar{\alpha}$, $p \varepsilon \bar{H}_{\bar{\alpha}}$. which is an $\mathfrak{A}$-block. Thus A satisfies * and **. We have shown that the sets satisfying * and ** form a Borel field including $\mathfrak{E}$. Since $\mathfrak{B}$ is included in any Borel field including $\mathfrak{E}$, every set in $\mathfrak{B}$ satisfies * and **.

COROLLARY. *If $B \varepsilon \mathfrak{B}$, then B is the union (possibly non-countable) of $\mathfrak{A}$-blocks in B.*

We now introduce some notation associated with product spaces, largely following Kakutani [K]. Let $\Omega^* =$ the set of all Γ-sequences, $\omega^* = \{\omega^\gamma \mid \gamma \varepsilon \Gamma\}$ with $\omega^\gamma \varepsilon \Omega$ for each $\gamma \varepsilon \Gamma$. Let $B^{*\gamma} = \{\omega^* \mid \omega^\gamma \varepsilon B^\gamma\}$, $\mathfrak{B}^{*\gamma} = \{B^{*\gamma}\}$, $\mathfrak{A}^* = \{\mathfrak{B}^{*\gamma} \mid \gamma \varepsilon \Gamma\}$. Note: $\mathfrak{A}^*$ is a denumerably independent family of Borel fields of subsets of Ω^*. Let $E^*[H^*]$ represent an $\mathfrak{A}^*$-rectangle [block] and let $\mathfrak{E}^*[\mathfrak{B}^*]$ represent the [Borel] field generated by the $\mathfrak{A}^*$-rectangles. The previous lemma, theorem, and corollary apply to family $\mathfrak{A}^*$.

If with each $\mathfrak{B}^{*\gamma}$ we have m^γ a countably additive measure (all measures are countably additive in the sequel), where $m^\gamma(\Omega^*) = 1$, then it is known, (e. g. [K]) that there is a unique measure m^* on $\mathfrak{B}^*$ such that for each $\mathfrak{A}^*$-block $H^* = \bigcap B^{*\gamma}$, $m^*(H^*) = \prod m^\gamma(B^\gamma)$. We are now in a position to prove the

THEOREM (BANACH). *If $\mathfrak{A}$ is a denumerably independent family $\{\mathfrak{B}^\gamma \mid \gamma \varepsilon \Gamma\}$ of Borel fields of subsets of Ω and for each $\mathfrak{B}^\gamma$, $\gamma \varepsilon \Gamma$, there is a measure m^γ such that $m^\gamma(\Omega) = 1$, then there is a unique measure m on $\mathfrak{B}$ the Borel field generated by $\mathfrak{A}$ such that*

$$A \varepsilon \mathfrak{B}^\gamma \Longrightarrow m(A) = m^\gamma(A),$$

$$\text{(stochastic independence) } \mathfrak{A}\text{-block } H = \bigcap_\gamma B^\gamma \Longrightarrow mH = \prod_\gamma m^\gamma(B^\gamma).$$

Proof. Let ϕ be the transformation from Ω into Ω^* defined by $\phi(\omega) = \{\omega^\gamma \mid \gamma \varepsilon \Gamma\}$ where $\omega^\gamma = \omega$ for every γ; let D^* be the range of ϕ.

Banach's theorem can be proved by writing $m\phi^{-1}(A^*) = m^*A^*$—all that needs proof is that ϕ^{-1} (which is obviously a σ-homomorphism from $\mathfrak{B}^*$ onto $\mathfrak{B}$) is an isomorphism. For this in turn it is sufficient to prove that if $A^* \varepsilon \mathfrak{B}^*$ and $A^* \bigcap D^* = \Lambda$ then $A^* = \Lambda$. If A^* is an $\mathfrak{A}^*$ block, $A^* = \bigcap_{\gamma} B^{*\gamma}$, $B^{*\gamma} = \{\omega^* \mid \omega^\gamma \varepsilon B^\gamma\}$, then $\phi^{-1}(A^*) = \bigcap_{\gamma} B^\gamma$ and the desired conclusion follows from the assumed denumerable independence; the general case is now a consequence of the corollary.[3] Note that the σ-isomorphism between $\mathfrak{B}^*$ and $\mathfrak{B}$ is a set-theoretical proposition independent of measure assumptions.

INSTITUTE FOR ADVANCED STUDY.

BIBLIOGRAPHY.

[B] S. Banach, "On measures in independent fields," *Studia Mathematica*, vol. 10 (1948), pp. 159-177.

[K] S. Kakutani, "Notes on infinite product measure spaces, I," *Proceedings of the Imperial Academy at Tokyo*, vol. 19 (1943), pp. 148-151.

[SJ] Erik Sparre Andersen and Børge Jessen, "Some limit theorems on integrals in an abstract set," *Det Kgl. Danske Videnskabernes Selskab, Matematisk-Fysiske Meddelelser*, vol. 22, no. 14 (1946).

---

[3] The method for proving the σ-isomorphism between $\mathfrak{B}^*$ and $\mathfrak{B}$ is the referee's and effects a considerable simplification of the author's method, which was long and roundabout.

CERTAIN CLASSES OF SERIES TO SERIES TRANSFORMATION MATRICES.*

By P. Vermes.

---

1. In a recent paper[1] there were described series to series transformation matrices having the property that the corresponding sequence to sequence transformation matrix was a scalar multiple of the same matrix or of the same matrix without its first row or column. In the present paper the most general matrices are found which have this property, in each case as a solution of a partial difference equation.

In matrix notation, if $A \equiv (a_{nk})$ $(n, k = 0, 1, \cdots)$ is a series to series transformation matrix, the corresponding series to sequence transformation matrix G is given by $g_{nk} = a_{0k} + a_{1k} + \cdots + a_{nk}$, and the corresponding sequence to sequence transformation matrix F by $f_{nk} = g_{nk} - g_{n\,k+1}$, [ST (2.4), (2.15)].

If E, R and S are the matrices

$$(1.1)\quad E \equiv \begin{matrix} 0, 0, 0, \cdots \\ 1, 0, 0, \cdots \\ 0, 1, 0, \cdots \\ 0, 0, 1, \cdots \\ \cdot\ \cdot\ \cdot\ \cdots \end{matrix}, \quad R \equiv \begin{matrix} 1, 0, 0, 0, \cdots \\ 1, 1, 0, 0, \cdots \\ 1, 1, 1, 0, \cdots \\ 1, 1, 1, 1, \cdots \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdots \end{matrix}, \quad S \equiv \begin{matrix} 1, & 0, & 0, 0, \cdots \\ -1, & 1, & 0, 0, \cdots \\ 0, & -1, & 1, 0, \cdots \\ 0, & 0, & -1, 1, \cdots \\ \cdot & \cdot & \cdot\ \cdots \end{matrix};$$

so that

$$(1.2)\qquad R = I + E + E^2 + \cdots, \qquad S = I - E, \qquad RS = SR = I,$$

the correspondence between the matrices A and F can be expressed in the form

$$(1.3)\qquad F = (RA)S = R(AS) = RAS,$$

since each element of the product matrix is the sum of a finite number of elements.

---

* Received August 29, 1949.

[1] "Series to series transformations and analytic continuation by matrix methods," *American Journal of Mathematics*, vol. 71 (1949), pp. 541-562. This paper will be referred to as [ST].

If E' is the transpose of E, the matrix product AE is the matrix A without its first column, and the product $E'A$ is the matrix A without its first row. In what follows we shall find all solutions of the matrix equations

$$RXS = px, \qquad RXS = pXE, \qquad RXS = pE'X, \qquad RXS = pE'XE,$$

where p is an arbitrary fixed number, and R, S, E have the meanings given above. In every case the solution is a matrix $A(p)$ depending on the number p multiplied by a matrix, the elements of which are arbitrary to a certain degree. The matrices $A(p)$ are in the various cases equal or simply related to the series to series transformation matrices of the Taylor series and Laurent series continuation and of the Euler summation matrix, [ST Sections **3**, **4**, **5**].

2. The equation $RXS = pX$. Since R and S are row-finite, we have for every X $S(RXS) = (SR)(XS) = XS$, and $R(SX) = (RS)X = X$, hence equation **2** implies and is implied by

$$(2.1) \qquad XS = pSX.$$

It follows from (2. 1) that the elements x_{nk} satisfy

$$(2.2) \qquad x_{n\,k+1} = (1-p)x_{nk} + px_{n-1\,k} \qquad k = 0, 1, \cdots, \quad n = 1, 2, \cdots,$$

$$(2.3) \qquad x_{0\,k+1} = (1-p)x_{0k} \qquad k = 0, 1, \cdots.$$

Equation (2. 2) is the partial difference equation to be solved, and (2. 3) represents incomplete boundary conditions leaving the elements x_{n0} of the first column arbitrary. Taking $x_{00} = 1$, and all other arbitrary elements zero, we obtain step by step all elements of a matrix A which we regard as the *fundamental solution*

$$a_{nk} = \binom{k}{n} p^n (1-p)^{k-n} \text{ for } k \geq n; \; a_{nk} = 0 \text{ for } k < n.$$

Taking $x_{N0} = 1$, and all other arbitrary elements zero, where N is a positive integer, the same matrix results with N rows of zeros added. The fundamental solution is the matrix $A(p)$ of the Taylor series continuation [ST (3. 14)], and the matrix with N rows of zeros added can be expressed as $E^N A(p)$, where E is the diagonal vector defined in (1. 1). Hence giving arbitrary values d_n to the elements of the first column, the other elements can be found from (2. 2) and (2. 3) step by step, each of the d_n contributing linearly a matrix $d_n E^n A(p)$. The *general solution* is therefore the sum of all the contributions

$$X = DA(p), \tag{2.4}$$

where

$$D = d_0 I + d_1 E + d_2 E^2 + \cdots = \begin{matrix} d_0, & 0, & 0, & \cdots \\ d_1, & d_0, & 0, & \cdots \\ d_2, & d_1, & d_0, & \cdots \\ \cdot & \cdot & \cdot & \cdots \end{matrix}. \tag{2.5}$$

The matrix D, being a power series in E, commutes with every power series in E, in particular with R and S. Hence $RDS = RSD = D$, i. e. the corresponding sequence to sequence transformation matrix is the same matrix D. Regarded as a sequence to sequence transformation matrix, it is a "modified Nörlund matrix."[2] It is regular if and only if $\Sigma \mid d_k \mid$ is convergent and $\Sigma d_k = 1$. In this case it is regular both as a series to series or as a sequence to sequence transformation matrix.

The properties of the matrix $A(p)$ have been discussed in [ST **3**]. It is regular if and only if $0 \leq p \leq 1$. Another property follows easily from (2.1): Since (Ap), $A(q)$, S are column-finite, $A(p)A(q)S = qA(p)SA(q) = qp \cdot SA(p)A(q)$, so that $A(p)A(q)$ is *the* fundamental solution of the equation $XS = pqSX$, i. e. $A(p)A(q) = A(pq)$, as proved by direct calculation in [ST 3. V.].

3. The equation $RXS = pXE$. As shown in Section **2**, this equation is equivalent to

$$XS = pSXE, \tag{3.1}$$

which requires $x_{nk} - x_{n\,k+1} = p(-x_{n-1\,k+1} + x_{n\,k+1})$, and $x_{0k} - x_{0\,k+1} = px_{0\,k+1}$ for $k = 0, 1, 2, \cdots;\ n = 1, 2, \cdots$.

Excluding $p = -1$, and putting $1/(1+p) = t$, we obtain

$$x_{n\,k+1} = tx_{nk} + (1-t)x_{n-1\,k+1}, \text{ for } n = 1, 2, \cdots, \tag{3.2}$$

and

$$x_{0\,k+1} = tx_{0k}. \tag{3.3}$$

---

[2] G. Piranian, "Nörlund transformations with a bounded base," Office of Naval Research, Project M786, 1949.

The elements of the first column can be chosen arbitrarily, and then all other elements are determinate, and can be found step by step. Taking $x_{00}=1$, and all other arbitrary elements zero, we obtain the *fundamental solution*

$$a_{nk}=\binom{k+n-1}{n}(1-t)^n t^k,$$

which is the matrix $A(t)$ of the Laurent series continuation [ST (5.3)]. The *general solution* is then obtained, as in Section 2,

$$X=DA(t). \tag{3.4}$$

When $p=-1$, we obtain $x_{nk}=x_{n-1\ k+1}$ and $x_{0k}=0$, which gives the zero matrix, already included in (3.4).

4. The equation $RXS=pE'X$. This equation is equivalent to

$$XS=pSE'X, \tag{4.1}$$

which requires, $x_{nk}-x_{n\ k+1}=p(x_{n+1\ k}-x_{nk})$, and $x_{0k}-x_{0\ k+1}=px_{1k}$ for $k=0,1,2,\cdots;\ n=1,2,\cdots$.

Excluding $p=0$, and putting $-1/p=q$, we obtain

$$x_{n+1\ k}=(1-q)x_{nk}+qx_{n\ k+1}, \qquad n=1,2,\cdots, \tag{4.2}$$

and

$$x_{1k}=q(x_{0\ k+1}-x_{0k}). \tag{4.3}$$

Here the elements of the first row can be chosen arbitrarily, and the other elements then determined step by step. Taking $a_{0k}=1$, for a fixed k, and all other elements of the first row zero, (4.3) determines the values in the second row $a_{1k}=-q$, $a_{1\ k-1}=q$, all the other elements zero. The elements in the other rows are then obtained easily by considering separately the values contributed by a_{1k} and $a_{1\ k-1}$. A short calculation shows that the matrix thus obtained can be expressed as the matrix product $B(q)SC_k$, where $B(q)$ is the bordered Euler series to series summation matrix $\mathcal{E}(q)$: [ST(4.2)]

$$B(q)\equiv\begin{matrix} 1, & 0\ , & 0\ , & 0\ , & 0\ ,\cdots \\ 0, & q\ , & 0\ , & 0\ , & 0\ ,\cdots \\ 0, & q(1-q)\ , & q^2\ , & 0\ , & 0\ ,\cdots \\ 0, & q(1-q)^2, & 2q^2(1-q), & q^3\ , & 0\ ,\cdots \\ 0, & q(1-q)^3, & 3q^2(1-q)^2, & 3q^3(1-q), & q^4,\cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot\ \cdots \end{matrix}\ . \tag{4.4}$$

S is defined in (1. 2), and C_k is the cross-diagonal vector

$$(4.5)\qquad C_k \equiv \begin{array}{l} 0, 0, \cdots, 0, 0, 1, 0, 0, \cdots \\ 0, 0, \cdots, 0, 1, 0, 0, 0, \cdots \\ 0, 0, \cdots, 1, 0, 0, 0, 0, \cdots \\ \cdot\ \ \cdot\ \ \cdots, \cdot\ \ \cdot\ \ \cdot\ \ \cdot\ \ \cdots \\ 0, 1, \cdots, 0, 0, 0, 0, 0, \cdots \\ 1, 0, \cdots, 0, 0, 0, 0, 0, \cdots \\ \cdot\ \ \cdot\ \ \cdots, \cdot\ \ \cdot\ \ \cdot\ \ \cdot\ \ \cdots \end{array} \qquad c_{ni} \begin{cases} = 1 \text{ for } n + i = k, \\ = 0 \text{ otherwise.} \end{cases}$$

The *general solution* is therefore

$$(4.6)\qquad X = B(q)SC, \text{ where } C \equiv \begin{array}{l} c_0, c_1, c_2, \cdots \\ c_1, c_2, c_3, \cdots \\ c_2, c_3, c_4, \cdots \\ \cdot\ \ \cdot\ \ \cdot\ \ \cdots, \end{array}$$

$c_0, c_1, \cdots$ being arbitrary numbers. When $p = 0$, we obtain $x_{nk} = x_{n\,k+1}$ for every n and k, which is satisfied by any matrix with all its columns equal. This is not included in (4. 6).

5. The equation $RXS = pE'XE$. This is equivalent to

$$(5.1)\qquad XS = pSE'XE,$$

which requires, $x_{nk} - x_{n\,k+1} = p(x_{n+1\,k+1} - x_{n\,k+1})$, and $x_{0k} - x_{0\,k+1} = px_{1\,k+1}$ for $k = 0, 1, 2, \cdots$; $n = 1, 2, \cdots$.

Excluding $p = 0$, and putting $1/p = q$, we obtain

$$(5.2)\qquad x_{n+1\,k+1} = (1 - q)x_{n\,k+1} + qx_{nk}, \qquad n = 1, 2, \cdots,$$

and

$$(5.3)\qquad x_{1\,k+1} = q(x_{0k} - x_{0\,k+1}).$$

These equations leave the elements of the first row and first column arbitrary. Taking $x_{00} = 1$, all the other arbitrary elements zero, we obtain as fundamental solution the matrix $B(q)$ defined in (4. 4). If, for a fixed n, $x_{n0} = 1$, the other arbitrary elements being zero, we obtain, as in Section **2**, the matrix $E^nB(q)$, so that the contribution of all the elements of the first column is the matrix $DB(q)$, D being defined in (2. 5). If the only nonzero arbitrary element is $x_{0k} = 1$, for a fixed $k \geq 1$, we obtain, as in Section **4**, the matrix $B(q)SE'^{k}$. Hence the contribution of arbitrary elements $x_{01}=h_1$, $x_{02}=h_2,\cdots$ is *a* matrix H:

$$(5.4)\qquad H=h_1E'+h_2E'^2+\cdots=\begin{matrix}0, & h_1, & h_2, & h_3, & \cdots\\ 0, & 0, & h_1, & h_2, & \cdots\\ 0, & 0, & 0, & h_1, & \cdots\\ 0, & 0, & 0, & 0, & \cdots\\ \cdot & \cdot & \cdot & \cdot & \cdots\end{matrix}\;.$$

The *general solution*, for $p\neq 0$, is therefore

$$(5.5)\qquad X=DB(q)+B(q)SH.$$

When $p=0$, we obtain the same matrix as in Section **4**.

In the special case, when we take

$$d_n=q(1-q)^n,\quad n=0,1,2,\cdots,\qquad h_k=0,\quad k=1,2,\cdots,$$

so that $X=DB(q)$, we obtain the matrix $\mathcal{E}(q)$ of the Euler series to series transformation.

BIRKBECK COLLEGE, UNIVERSITY OF LONDON.

SUMMABILITY METHODS WEAKER THAN CONVERGENCE.*

By J. D. Hill.

Consider the Banach space (c) of all real convergent sequences $x \equiv \{s_k\}$ with $\| x \| \equiv \sup_k | s_k |$. Every linear functional $f(x)$ defined in (c) is of the form [1, p. 65] $f(x) = C \lim_k s_k + \sum_{k=1}^{\infty} C_k s_k$ with $\| f \| = | C | + \sum_{k=1}^{\infty} | C_k |$. A sequence

$$f_n(x) = C_n \lim_k s_k + \sum_{k=1}^{\infty} C_{nk} s_k \qquad (n = 1, 2, 3, \cdots) \tag{1}$$

of such functionals, in which we shall suppose that *not all of the* C_n *are zero,* defines a method of summability by means of which a sequence x is said to be summable to the value $f(x)$ if $\lim_n f_n(x)$ exists and equals $f(x)$. We shall refer to such a method as a *W-method.* If $f(x)$ is defined for every x in (c) then W is called *conservative.* It is a trivial extension of a standard theorem for Silverman-Toeplitz methods, and may be proved very simply by operational means [see 1, pp. 90-91], that W will be conservative if and only if the following conditions hold:

$$\lim_{n \to \infty} C_{nk} \equiv L_k \text{ exists}, \qquad (k = 1, 2, 3, \cdots), \tag{2}$$

$$\lim_{n \to \infty} \{C_n + \sum_{k=1}^{\infty} C_{nk}\} \equiv L \text{ exists}, \tag{3}$$

$$\sup_n \{| C_n | + \sum_{k=1}^{\infty} | C_{nk} |\} < \infty. \tag{4}$$

On the other hand, if $\lim_n f_n(x)$ exists and equals $\lim_k s_k$ for every x in (c), then W is *regular.* The conditions (2), (3), and (4) with all $L_k = 0$ and $L = 1$ are necessary and sufficient for regularity.

A transformation of the type W differs from the more familiar Silverman-Toeplitz matrix method in that the existence of $\lim_k s_k$ is required. Hence the *convergence-field* of W (i. e., the set of all W-summable sequences) is a subset of (c). Such a method is said to be *not stronger than convergence.* If the convergence-field is a *proper* subset of (c) then the method is *weaker than convergence.* The Cesàro methods (C, α) for $-1 < \alpha < 0$ are well known examples of Silverman-Toeplitz methods having this property [see 3, p. 17, Theorem II].

The following question arises. What properties must be possessed by a

---

* Received January 25, 1950.

proper subset G of (c) in order that there exist a W-method whose convergence-field is precisely G? It is first of all necessary that G be linear and Borel measurable [1, p. 18, Theorem 9] and consequently of the first category [1, p. 36, Theorem 1]. Closer to the heart of the matter is the fact that G must be a set of type $F_{\sigma\delta}$ [1, p. 18, Theorem 9], namely, a denumerable product of sets each of which is a denumerable sum of closed sets. It is known [1, p. 235] that if G is not closed then it cannot be a set of type F_σ (a denumerable sum of closed sets). The object of this note is to point out that every linear closed set is a set G. Whether every linear $F_{\sigma\delta}$ is a set G is a more difficult question which we propose to make the subject of a later paper. It will be observed from the proof that the following theorem can be rephrased to hold for sequences of linear functionals defined over any separable Banach space. For the present purposes, however, we prefer the form given.

THEOREM. *Let G be a linear proper subset of (c). Then in order that there exist a W-method whose convergence-field is precisely G it is sufficient (but not necessary) that G be closed.*

Proof. To see first that the condition is not necessary, let $X \equiv \{\delta_k{}^k\}$ and $X_n \equiv \{\delta_k{}^n\}$ for $n = 1, 2, 3, \cdots$, where $\delta_k{}^n$ is the Kronecker symbol. Let E denote the (proper) subset of (c) composed of all finite linear combinations of X and the X_n, and consider the W-transformation

$$(5) \qquad f_n(x) = \tfrac{1}{2}\lim_k s_k + (2\log 2)^{-1}\sum_{k=1}^{n} (-1)^{n-k} s_k/(n-k+1) \qquad (n = 1, 2, 3, \cdots).$$

The method (5) satisfies conditions (2) and (3) but not (4). Hence the convergence-field G of (5) contains E but is a proper subset of (c). It follows that G is not closed since the closure of E is (c) itself [1, p. 65, §3], and G contains E.

Now let G denote any closed linear proper subset of (c) and G^* its complement with respect to (c). Since (c) is a separable space, its subset G^* is likewise separable [3, p. 31], and we designate by $\mathfrak{Y}_1, \mathfrak{Y}_2, \mathfrak{Y}_3, \cdots$ a dense enumerable subset of G^*. Inasmuch as G is closed, the distance $d(y)$ from G to y is greater than zero for each y in G^*. In particular we have $d(\mathfrak{Y}_n) > 0$ for all n. Hence [1, p. 57, Lemma] for each $n = 1, 2, 3, \cdots$ there exist linear functionals $F_n(x)$ defined in (c) such that (i) $F_n(\mathfrak{Y}_n) = 1$; (ii) $F_n(x) = 0$ for all x in G; and (iii) $\| F_n \| = 1/d(\mathfrak{Y}_n)$. Let $Q_n(x) \equiv F_n(x)/\| F_n \|$ so that $\| Q_n \| = 1$ $(n = 1, 2, 3, \cdots)$ and let $\{F^*{}_n\}$ denote the sequence

$$(6) \qquad Q_1;\ -Q_1, -Q_2;\ Q_1, Q_2, Q_3;\ -Q_1, -Q_2, -Q_3, -Q_4;\ \cdots$$

in which the k-th group, set off by semicolons, contains the first k of the Q_n with the sign of $(-1)^{k-1}$.

It is first of all evident from (ii) that the sequence (6) applied to each x_0 in G yields a sequence composed wholly of zeros. We show next that the sequence (6) applied to each y_0 in G^* yields a divergent sequence bounded by $\| y_0 \|$. The boundedness follows directly from the fact that $\| Q_n \| = 1$ for all n. To establish the divergence we recall that $d(y_0) > 0$ and hence that we can fix a point $\mathfrak{Y}_p$ of the dense enumerable subset of G^* so that $\| y_0 - \mathfrak{Y}_p \| < \frac{1}{3}d(y_0)$. It is immediate then that $d(\mathfrak{Y}_p) \geqq \frac{2}{3}d(y_0)$. Consider the equation $F^*_n(y_0) = F^*_n(y_0 - \mathfrak{Y}_p) + F^*_n(\mathfrak{Y}_p)$. From (6), (i), (iii), and the definition of $Q_n(x)$, there exists a sequence of indices $n_k \uparrow \infty$ for which we have $F^*_{n_k}(\mathfrak{Y}_p) = (-1)^k Q_p(\mathfrak{Y}_p) = (-1)^k d(\mathfrak{Y}_p)$; consequently $F^*_{n_k}(y_0) = F^*_{n_k}(y_0 - \mathfrak{Y}_p) + (-1)^k d(\mathfrak{Y}_p)$. The first term on the right does not exceed $\frac{1}{3}d(y_0)$ in absolute value while the second alternates in sign and is never numerically less than $\frac{2}{3}d(y_0)$. This establishes the divergence of the sequence $\{F^*_n(y_0)\}$ for every y_0 in G^*.

Suppose now that $F^*_n(x) = C^*_n \lim_k s_k + \sum_{k=1}^{\infty} C^*_{nk} s_k$ and that $f_n(x)$, given by (1), defines an arbitrary conservative (or regular) W-method such that not all of the numbers $C^*_n + C_n$ are zero. We set

$$(7) \qquad g_n(x) = F^*_n(x) + f_n(x) \text{ for } x \text{ in } (c), \qquad (n = 1, 2, 3, \cdots).$$

It is apparent from what precedes that (7) defines a method W^* which is effective (or regularly effective) in G and definitely ineffective in G^*. This completes the proof.

It may be observed incidentally that the norm condition (4) is satisfied by the method (7). However, there exist W-methods having the required property and for which (4) is not satisfied; for example, the method defined by $h_n(x) = \lambda_n F^*_n(x) + f_n(x)$ where $\lambda_n \to \infty$.

MICHIGAN STATE COLLEGE.

REFERENCES.

[1] S. Banach, *Théorie des opérations linéaires*, Warsaw, 1932.

[2] E. Kogbetliantz, *Sommation des séries et intégrales divergentes par les moyennes arithmétiques et typiques*, Paris, 1931.

[3] M. H. A. Newman, *Elements of the topology of plane sets of points*, Cambridge, 1939.

ERRATA.

W. L. Chow, "On compact complex analytic varieties," this JOURNAL, vol. 71, pp. 893-914.

Page 894, line 8 from the bottom, "The set W_r . . ." should read "The set $\tilde{W}_r$ . . ."

Page 894, line 5 from the bottom, "The set $W_r — W_r$ of . . ." should read "The set $W_r — W_r$ of . . ."

Page 895, line 8, "analytic elements . . ." should read "regular analytic elements . . ."

Philip Hartman and Aurel Wintner, "On linear difference equations of second order," this JOURNAL, vol. 72, pp. 124-128.

Page 124, in formula (3), replace the second sign "$>$" by "$<$."

Page 125, line 14, replace "non-decreasing" by "non-increasing."

Page 125, in the first formula line, replace "$2p_k$" by "2."

W. L. Chow, "Algebraic systems of positive cycles in an algebraic variety," this JOURNAL, vol. 72, pp. 247-283.

Page 251, line 18, "Let K be . . ." should read "let $\breve{K}$ be . . ."

Page 251, line 20, ". . . over K is called the extension U/K of U/K over K." should read ". . . over $\breve{K}$ is called the extension $U/\breve{K}$ of U/K over $\breve{K}$."

Page 251, line 25, ". . . an extension W/K over . . ." should read ". . . an extension $W/\breve{K}$ over . . ."

Page 281, line 16, ". . . in this case the $\mathfrak{m}$-adic ring . . ." should read ". . . in this case the $(L \times \mathfrak{o}^*)\mathfrak{m}$-adic ring . . ."

FREE SUMS OF GROUPS AND THEIR GENERALIZATIONS II.*

By Reinhold Baer.

Recently we have shown that the appropriate framework for a discussion of the existence of free sums with amalgamations is provided by the concept of add and that one has to determine those adds which are amalgams of their subgroups and satisfy the strong associative law. Since the latter law is stated in terms of vectors over the add, the problem arises of giving an intrinsic characterization; and this problem we are going to solve in the present investigation [Section **4**]. An application of this result is then given in Section **5**.

For our purposes [and also for later applications] it proves convenient to analyze the concept of amalgam of subgroups and to break it down into formal laws concerning the addition of elements. These formal laws are the "postulates" of Section **1**; and in Sections **2** and **3** we show that they really characterize amalgams of subgroups. In order to obtain these results we have to solve the problem [of interest in itself] to characterize those subsets of an add which are contained in a subgroup of the add [Section **2**].

1. The postulates. An *add* is a set A of elements subject to a composition which we call addition and which meets only the following requirement:

The sum of the elements a and b in A is either undefined or else it is a uniquely determined element $a + b$ *in* A.

We note the possibility that the sum $a + b$ is defined while the sum $b + a$ is undefined.

The adds we shall be interested in will always be of a more special type. We shall therefore enumerate a number of formal requirements. These postulates will not be independent. But we shall require at various stages of our investigation the validity of various sets of these postulates; and thus this abundance will prove a convenience.

Postulates of Inversion.

I. *There exists an element* 0 *in* A *such that* $x + 0 = 0 + x = x$ *for every element* x *in* A.

It is clear that 0 is uniquely determined by Postulate I.

---

* Received June 6, 1949.

II. *To every x in A there exists y in A such that* $x + y = 0$.

III. $x + y = 0$ *and* $x + z = 0$ *imply* $y = z$.

If I, II and III are satisfied by A, then there exists to every element x in A one and only one element $-x$ in A such that $x + (-x) = 0$. This symbol $-x$ will be used whenever we are assured of the validity of Postulates I, II, III; and then we shall write $x - y$ instead of $x + (-y)$ etc.

IV. $x + y = 0$ *implies* $y + x = 0$.

Postulates I to IV imply $-(-x) = x$. Adds which meet these four requirements have been termed elsewhere self-reflexive [Baer (1), Section **4**].

V. $x + (y + z) = 0$ *if, and only if,* $(x + y) + z = 0$.

This is in reality a rather elaborate statement. It asserts that the existence of the sums $y + z$ and $x + (y + z)$ and the validity of the equation $x + (y + z) = 0$ are equivalent to the existence of the sums $x + y$ and $(x + y) + z$ together with $(x + y) + z = 0$. The following proposition will be useful.

(0) *If Postulates* I *and* IV *are satisfied, then the following properties of the add A are equivalent.*

(i) *Postulate* V.

(ii) $x + (y + z) = 0$ *implies* $(x + y) + z = 0$.

(iii) $(x + y) + z = 0$ *implies* $x + (y + z) = 0$.

Proof. It clearly suffices to show that (ii) implies (iii). Hence assume the validity of (ii) and suppose that $(x + y) + z = 0$. Then we deduce from IV and (ii) successively the validity of the following equations: $z + (x + y) = 0$, $(z + x) + y = 0$, $y + (z + x) = 0$, $(y + z) + x = 0$, $x + (y + z) = 0$, so that (iii) is true.

POSTULATES OF ASSOCIATIVITY.

VI. *If the sums* $x + y$, $y + z$, $(x + y) + z$, $x + (y + z)$ *exist in A, then* $(x + y) + z = x + (y + z)$.

VII. *If the sums* $x + y$ *and* $y + z$ *exist in A, then the existence of* $(x + y) + z$ *is necessary and sufficient for the existence of* $x + (y + z)$.

It is worth noting that the antecedents of the postulates V, VI, VII require the existence of 2, 4, 3 sums respectively. We note the validity of the following important rule.

(1) *Assume validity of* I, VI, VII *and existence of* $r+s$. *Then* $q+r=0$ *implies* $q+(r+s)=s$ *and* $s+t=0$ *implies* $(r+s)+t=r$.

This is an immediate consequence of $s=(q+r)+s$ and $r=r+(s+t)$.

It follows from (1) that adds satisfying Postulates I to VII are self-reflexive in the strict sense [Baer (1), Section **4**].

(2) *If Postulates* I, V, VI, VII *are satisfied by* A, *if* $x+y$ *exists in* A, *and if* $x+x'=y+y'=0$, *then* $y'+x'$ *exists in* A *and satisfies* $(x+y)+(y'+x')=0$.

Proof. We infer from (1) that $x=(x+y)+y'$. Consequently $0=[(x+y)+y']+x'$; and our contention is a consequence of *V*.

(3) *Postulates* III, IV *are consequences of* I, II, VI.

Proof. Assume $x+y=0$. There exists by II an element z such that $y+z=0$. It follows from I that $x=x+(y+z)$, $z=(x+y)+z$; and consequently we may infer from VI that $x=x+(y+z)=(x+y)+z=z$. Thus IV is a consequence of I, II, VI.

If $x+y=x+z=0$, then we infer from IV that $z+x=0$; and it follows as before that $y=(z+x)+y=z+(x+y)=z$. Hence III is a consequence of I, IV, VI and therefore of I, II, VI.

(4) *Postulate* V *is a consequence of* I, II, VI, VII.

Proof. Assume that $x+(y+z)=0$. There exists by II an element w in A such that $z+w=0$. Hence it follows from (1) that $y=(y+z)+w$; and we have by I that $w=[x+(y+z)]+w$. Now it follows from VI, VII that

$$w=[x+(y+z)]+w=x+[(y+z)+w]=x+y.$$

From IV and $z+w=0$ we infer now that $0=w+z=(x+y)+z$. Thus we have shown that $x+(y+z)=0$ implies $(x+y)+z=0$.

From this fact and IV we deduce successively: if $(r+s)+t=0$, then $t+(r+s)=0$, $(t+r)+s=0$, $s+(t+r)=0$, $(s+t)+r=0$, $r+(s+t)=0$; and this completes the proof of the validity of V.

We notice that all the postulates I to VII are valid whenever we are assured of the validity of I, II, VI, VII. We shall make much use of this fact.

POSTULATES OF EXISTENCE.

VIII. *The existence of the sums* $x+y$, $y+z$, $x+z$ *implies the existence of* $(x+y)+z$.

In conjunction with VII this implies also the existence of $x+(y+z)$.

(5) *If Postulates* I, V, VI, VII, VIII *are satisfied by A, if the sums $x+x$, $x+y$, $y+y$ exist in A, and if there exist elements x', y' such that $x+x'=x'+x=y+y'=y'+y=0$, then the sums $x+y'$ and $y+x$ exist in A.*

Proof. We infer from (2) the existence of $y'+y'$ and from (1) the validity of $x=(x+y)+y'$. From VIII and the existence of the sums $(x+y)+y'$, $y'+y'$, $(x+y)+y'$ we infer the existence of

$$[(x+y)+y']+y'=x+y'.$$

From $x+x'=y'+y=0$ and the existence of $x+y'$, just established, we infer by (2) the existence of $y+x'$. But then it follows from our hypotheses and the result of the first paragraph that $y+x$ exists too, as we desired to show.

IX. *The sum $x+y$ exists in A if, and only if, $y+x$ exists in A.*

X. *The sum $x+x$ exists for every x in A.*

(6) *Postulate* IX *is a consequence of Postulates* I, II, VI, VII, VIII *and* X.

This is a fairly immediate consequence of (3), (4), (5).

We shall show in the next section that these last three postulates assure the existence of sufficiently many subgroups of the add A; and this is the reason for terming these postulates the Postulates of Existence. We note that these last three postulates have considerably more depth than the preceding ones. It is for this reason that we shall try to avoid using them as much as possible.

2. The subgroups of an add. Throughout this section we shall consider adds satisfying Postulate I, so that we may speak always of the zero element in A. Every subset of the add A may be considered as a subadd of A. If a subadd of A contains 0 and is a group with respect to the addition defined in A, then we shall term it *a subgroup of A*. Our main problem in this section is the determination of those subsets of A which are contained in subgroups of A. With this in mind we state

DEFINITION 1. *The subset S of the add A is a C-set, if*

(C. 1) *the sum $x+y$ exists in A, whenever x and y are in S; and*

(C. 2) *to every s in S there exists an element s' in S such that $s+s'=0$.*

Proposition 1. *If Postulates* I, VI, VII, VIII *are satisfied by the add A, then every C-set is contained in a subgroup of A.*

Proof. Suppose that S is a C-set in the add A. Denote by $S(0)$ the subset of A which consists of 0 and the elements in S. Clearly $S(0)$ is a C-set too. We define now inductively subsets $S(i)$ of A by the following rule:

(0) $S(i+1) = S(i) + S(i)$ consists of all the sums $x+y$ for x, y in $S(i)$ which exist in A.

We prove now inductively:

(1) $S \leq S(0) \leq S(1) \leq \cdots \leq S(i) \leq S(i+1) \leq \cdots$.

(2) Every $S(i)$ is a C-set.

The assertion (1) is an immediate consequence of the fact that $T \leq T+T$, provided 0 is in T. To prove (2) we note first that $S(0)$ is a C-set. Hence we may make the inductive hypothesis that $S(i)$ is a C-set. If x and y are elements in $S(i)$, then their sum $x+y$ exists in A [by (C. 1)] and belongs to $S(i+1)$. We infer from (C. 2) the existence of elements x', y' in $S(i)$ such that $x+x' = y+y' = 0$. It follows from (C. 1) that the sum $y'+x'$ of the elements y' and x' in $S(i)$ exists in A and belongs to $S(i+1)$. We infer from § 1, (1) that $x = (x+y)+y'$. Hence $0 = [(x+y)+y']+x'$. Now we infer from the existence of $y'+x'$ in A and from Postulates VI, VII that

$$0 = [(x+y)+y']+x' = (x+y)+(y'+x');$$

and this shows the validity of (C. 2) in $S(i+1)$. If x, y, z are elements in $S(i)$, then we infer from (C. 1) the existence of the sums $x+y$, $y+z$, $x+z$ in A; and now we may infer from Postulate VIII the existence of $(x+y)+z$ in A. If a, b, c, d are elements in $S(i)$, then we infer from the preceding remark and from (C. 1) the existence of the sums $(a+b)+c$, $c+d$, $(a+b)+d$ in A; and consequently we may deduce from Postulate VIII the existence of $[(a+b)+c]+d$. But $c+d$ exists in A [by (C. 1)]; and now we infer from Postulate VII the existence of the sum $(a+b)+(c+d)$ in A. This shows the validity of (C. 1) in $S(i+1)$. Hence $S(i+1)$ is a C-set too; and this completes the inductive proof of (2).

Denote by G the set-theoretical join of the sets $S(i)$. It follows from (1) that $S \leq G$. If x and y are elements in G, then there exists by (1) an integer m such that x and y are both in $S(m)$. It follows from (2) that $S(m)$ is a C-set so that the sum $x+y$ exists in A. Hence $x+y$ is in

$S(m+1)$ by (0): and consequently $x+y$ is in G. Thus we see that the sum of elements in G is a uniquely determined element in G; and it follows from Postulate VI that addition in G is associative. From $S(0) \leq G$ we infer that 0 is in G. If finally x belongs to G, then x belongs to some $S(k)$; and it follows from (2) that $S(k)$ contains an element x' such that $x+x'=0$. Hence there exists to every x in G an element x' in G such that $x+x'=0$; and now it is easily verified that G is a subgroup of the add A which contains S, as we intended to prove.

PROPOSITION 2. *If the Postulates* I, V, VI, VII, VIII *are satisfied by the add A, then the following conditions are necessary and sufficient for the subset S of A to be part of a subgroup of the add A.*

(a) *If x and y are in S, then the sum $x+y$ exists in A.*

(b) *To every x in S there exists an element x' in A such that $x+x'=x'+x=0$.*

Proof. The necessity of the conditions (a) and (b) is obvious. Assume now that the subset S of A satisfies the conditions (a) and (b). Then we may select to every element s in S some element s' in A such that $s+s'=s'+s=0$; and we may form the set T of all the elements s and s' for s in S. It is clear that Condition (C. 2) is satisfied by T. If x and y are elements in S, then we infer from Condition (a) the existence of the sums $x+x$, $x+y$, $y+y$, $y+x$ in A. Hence we may deduce from §1, (2) and §1, (5) the existence of the sums $x'+x'$, $y'+x'$, $y'+y'$, $y'+x'$ and $x+y'$, $y'+x$. This proves the validity of (C. 1) in T. Hence T is a C-set; and it follows from Proposition 1 that T is contained in a subgroup of the add A. But S is part of T; and so S itself is part of a subgroup of the add A.

COROLLARY 1. *Assume the validity of the postulates* I, II, VI, VII, VIII *in the add A. Then*

(a) *the subset S of A is part of a subgroup of A if, and only if, $x+y$ exists in A whenever x and y are in S.*

(b) *Every subgroup of A is part of a maximal subgroup of A.*

(c) *If M is a maximal subgroup of A, if w is an element in A, but not in M such that $w+w$ exists in A, then there exists an element v in M such that at least one of the sums $w+v$ and $v+w$ does not exist in A.*

Proof. We infer from §1, (3) and §1, (4) the validity of all the postulates I to VIII in A. Now (a) is an immediate consequence of Proposition 2.

The contention (b) is verified by the customary set-theoretical arguments [Maximum Principle]. Suppose now that G is a subgroup of the add A, and that w is an element in A such that the sums $w+w$, $w+g$, $g+w$ exist for every g in G. If S is the set composed of w and the elements in G, then the sum $x+y$ exists in A whenever x and y are in S. It follows from (a) that S is part of a subgroup of A. If G were in particular a maximal subgroup of A, then this would imply that w is an element in G; and this proves (c).

DEFINITION 2. *The add A is the amalgam of its subgroups [in the strict sense], if*

(A. 1) *every element in A is contained in a subgroup of A;*

(A. 2) *cross cuts of subgroups of A are subgroups of A;*

(A. 3) *the existence of the sums $x+y$, $y+z$, $z+x$ in A implies the existence of a subgroup of A which contains x, y, z.*

Amalgams in the weak sense will be defined in the next section. [See also Baer (1), §**6**.]

COROLLARY 2. *If the Postulates* I, II, VI, VII, VIII *are satisfied by the add A, and if the subadd $E = E(A)$ consists of all the elements e in A such that the sum $e+e$ exists in A, then $E(A)$ has the following properties.*

(a) *If x is in E, and if x' is an element in A such that $x+x'=0$, then x' belongs to E.*

(b) *If x and y are in E, and if the sum $x+y$ exists in A, then $x+y$ belongs to E.*

(c) *Every subgroup of A is part of E.*

(d) *E is the amalgam of its subgroups [in the strict sense].*

Proof. We note first that [by §**1**, (3) and §**1**, (4)] all Postulates I to VIII are satisfied by A. If $x+x$ exists in A, and if $x+x'=0$, then it follows from §**1**, (2) that $x'+x'$ exists in A; and this proves (a). If $x+x$, $x+y$, $y+y$ exist in A, then we infer from §**1**, (5) the existence of the sum $y+x$ in A. It follows from Corollary 1, (a) that there exists a subgroup of A which contains x and y; and this implies in particular the existence of the sum $(x+y)+(x+y)$. This proves (b). (c) is trivial. It is a consequence of Corollary 1, (a) that the element x in A belongs to a subgroup of A if, and only if, $x+x$ exists in A. Hence (A. 1) is satisfied by E.

That (A. 2) is satisfied by E is a consequence of the fact that the equation $a + x = 0$ possesses for every a in A one and only one solution x in A. Consider finally elements x, y, z in E such that the sums $x + y$, $y + z$, $z + x$ exist in A. Then it follows from §1, (5) that the sums $y + x$, $z + y$, $x + z$ exist in A; and the sums $x + x$, $y + y$, $z + z$ exist in A, since x, y, z are in E. Hence it follows from Corollary 1, (a) that there exists a subgroup S of A which contains x, y, z; and that $S \leq E$, is obvious [and has been pointed out before]. Thus E is the amalgam of its subgroups, as we claimed.

THEOREM. *The add A is the amalgam of its subgroups [in the strict sense] if, and only if, Postulates* I, II, VI, VII, VIII, X *are satisfied by A.*

Proof. It is a consequence of Corollary 2, (d) that A is an amalgam, if Postulates I, II, VI, VII, VIII, X are satisfied by A, since X implies $A = E(A)$. Assume conversely that A is the amalgam of its subgroups [in the strict sense]. It is a consequence of (A. 2) that all subgroups of A have the same null-element; and now we deduce Postulate I from (A. 1). Postulates II and X are likewise consequences of (A. 1). If the sum $x + y$ exists in A, then we infer from (A. 3) and the existence of the sums $0 + x$, $x + y$, $y + 0$ that $0, x, y$ are contained in some subgroup of A. Thus we have verified:

(A. 3*) *The sum $x + y$ exists in A if, and only if, x and y are contained in some subgroup of A.*

If now the sums $x + y$, $y + z$, $z + x$ exist in A, then it follows from (A. 3*) that the sum $x + z$ exists likewise. It follows from (A. 3) that x, y, z are contained in some subgroup of A; and thus $(x + y) + z$ exists in A. This shows the validity of VIII.

If the sums $x + y$, $y + z$, $(x + y) + z$ exist in A, then it follows from (A. 3*) that x, y belong to one subgroup and that y, z belong to some subgroup of A. Thus we are assured of the existence of the sums $y + (x + y)$, $(x + y) + z$, $z + y$; and we infer from (A. 3) the existence of a subgroup S of A which contains y, $x + y$, z. But it is clear that x, y, z belong to the subgroup S of A. Now one sees readily the validity of the following property.

(A. 3**) *The following properties of the elements x, y, z in the amalgam A are equivalent.*

(i) *There exists a subgroup of A which contains x, y, z.*

(ii) *The sums $x + y$, $y + z$, $z + x$ exist in A.*

(iii) *The sums* $x + y$, $y + z$, $(x + y) + z$ *exist in* A.

(iv) *The sums* $x + y$, $y + z$, $x + (y + z)$ *exist in* A.

From (A. 3**) one deduces readily the validity of Postulates VI and VII. This completes the proof.

Remark. It is a consequence of §1, (3), (4), (6) that all Postulates I to X are satisfied by amalgams.

3. Amalgams satisfying the strong associative law. The characterization of amalgams which we obtained in the preceding section can be simplified considerably, if we assume that the add under consideration satisfies the strong associative law. In order to state this law in the form best suited for our purposes some concepts have to be recalled.

If $a_1, \cdots, a_n$, $0 < n$, are elements in the add A, then we term the ordered n-tuplet $(a_1, \cdots, a_n)$ a vector of length n over A [and sometimes it will simplify notations to make use of the vector of length 0 too which is the empty set]. If $(a_1, \cdots, a_n)$ is a vector a over A, and if the sum $a_i + a_{i+1}$ exists in A, then we term the vector $(a_1, \cdots, a_{i-1}, a_i + a_{i+1}, a_{i+2}, \cdots, a_n)$ of length $n - 1$ a *contraction* of the vector a of length n. If the vector b arises from the vector a by a sequence of contractions, then b is termed a *reduction* of a [note that every vector is its own reduction too]. Finally we say that the vectors u and v are similar $[u \sim v]$, if there exist vectors $w(0), \cdots, w(i), \cdots, w(2k)$ such that $u = u(0)$, $w(2k) = v$ and such that $w(2i)$ and $w(2i + 2)$ are reductions of $w(2i + 1)$. [For a discussion of the significance of these concepts see Baer (1); but note that we have slightly changed the terminology.]

SIMILARITY CRITERION. *If Postulates* I *to* VII *are satisfied by the add* A, *then similarity of the vectors* u *and* v *is a necessary and sufficient condition for the existence of a vector* w *such that* u *and* v *are reductions of* w.

For a proof see Baer (1), §4, Corollary 1, where for Postulates I to VII the slightly weaker hypothesis has been substituted that A be strictly self-reflexive.

THE STRONG ASSOCIATIVE LAW *holds in the add* A, *if*

(S. 1) $(a) \sim (b)$ *implies* $a = b$ *and*

(S. 2) *similarity of the irreducible vector* v *to a vector of length* 1 *implies that* v *too be of length* 1.

Here we term a vector irreducible, if it does not possess any reductions. It has been shown [Baer (1), §3, Theorem 1] that this formulation of the strong associative law is equivalent with the one originally proposed [see Baer (1) for a discussion of the importance of this law.]

LEMMA 1. *The strong associative law implies Postulates* VI *and* VII.

Proof. If the sums $x + y$, $y + z$ and $(x + y) + z$ exist in A, then we have the vector similarities

$$((x + y) + z) \sim (x + y, z) \sim (x, y, z) \sim (x, y + z).$$

Now it follows from (S. 2) that the last vector is reducible. Hence there exists the sum $x + (y + z)$ in A. Consequently the vectors $((x + y) + z)$ and $(x + (y + z))$ are similar over A; and we infer $(x + y) + z = x + (y + z)$ from (S. 1). If the sums $x + y$, $y + z$ and $x + (y + z)$ exist in A, then we show similarly the existence of $(x + y) + z$ and the equality $x + (y + z) = (x + y) + z$. This shows the validity of VI and VII.

DEFINITION. *The add A is the amalgam of its subgroups [in the weak sense], if*

(B. 1) *every element in A belongs to some subgroup of A;*

(B. 2) *cross cuts of subgroups of A are subgroups of A;*

(B. 3) *the existence of the sum $x + y$ in A implies the existence of a subgroup of A which contains both x and y.*

LEMMA 2. *The add A is an amalgam in the strict sense if, and only if,*

(a) *A is an amalgam in the weak sense and*

(b) *to any given subgroups X, Y, Z of A there exists a subgroup S of A which contains the cross cuts $X \cap Y$, $Y \cap Z$, $Z \cap X$.*

Proof. Assume that A is an amalgam in the strict sense. Then it is easily seen that A is an amalgam in the weak sense too [we have verified this in §2, when we deduced (A. 3*)]. Consider now subgroups X, Y, Z of the add A and element r, s which belong to the join of $X \cap Y$, $Y \cap Z$ and $Z \cap X$. Then r and s belong to one of the subgroups X, Y, Z; and thus their sum $r + s$ exists in A. Since in A all Postulates I to X are satisfied [§2, Theorem], we may apply §2, Corollary 1, (a); and this proves the existence of a subgroup S of A which contains $X \cap Y$, $Y \cap Z$, $Z \cap X$. Thus the conditions (a) and (b) are necessary. If conversely the conditions (a) and (b) are satisfied by A, then A certainly satisfies (A. 1), (A. 2). If x, y, z are elements such that the sums $x+y$, $y+z$, $z+x$ exist in A, then we infer from (a) and (B. 3) the existence of subgroups X, Y, Z of A such that x, y belongs to X, y, z to Y and z, x to Z. But then y is in $X \cap Y$, z in $Y \cap Z$ and x in $Z \cap X$. It follows from (b) that there exists a subgroup S of A which contains these cross cuts and which therefore contains the elements x, y, z. Thus (A. 3) is satisfied by A too; and this completes the proof.

On the basis of this lemma it is easy to construct amalgams in the weak sense which are not amalgams in the strict sense.

THEOREM. *If the strong associative law is satisfied by the add A, then the following properties of A are equivalent.*

(i) *A is an amalgam in the strict sense.*

(ii) *A is an amalgam in the weak sense.*

(iii) *Postulates* I, II, IX, X *are satisfied by A.*

(iv) *Postulates* I, II, VIII, X *are satisfied by A.*

Proof. We have mentioned before [Lemma 2] that (i) implies (ii): and that amalgams in the weak sense satisfy Postulates I, II, IX, X, is easily deduced from the definition.

Assume now the validity of Postulates I, II, IX, X. We infer from Lemma I [and the strong associative law] the validity of Postulates VI, VII so that we may deduce from §1, (3), (4) the validity of Postulates III, IV, V. Thus we may use the symbol $-x$ in the usual fashion. We prove first:

IX*. *The existence of $x+y$ implies the existence of $x-y$.*

If $x+y$ exists in A, then there exist the sums $y+x$, $-y-y$, $x+x$. Consequently the following vector similarities are true:

$$(x+y, -y-y, y+x) \sim (x, y, -y, -y, y, x) \sim (x, x) \sim (x+x).$$

This implies by (S. 2) the reducibility of the first of these vectors; and consequently at least one of the sums $(x+y)+(-y-y)$ and $(-y-y)+(y+x)$ exists in A. From §1, (1) we infer $x=(x+y)-y$ and $x=-y+(y+x)$; and now we deduce from VI, VII the existence of $(x+y)+(-y-y)=[(x+y)-y]=y=x-y$ or of $(-y-y)+(y+x)=-y+[-y+(y+x)]=-y+x$. But it follows from IX that $x - y$ exists if, and only if, $-y + x$ exists; and this completes the proof of IX*.

Now we are ready to verify VIII. Assume the existence of the sums $x + y$, $y + z$, $x + z$ in A. This implies [by IX*] the existence of $x - y$, $y - z$, $x - z$. Consequently we have the vector similarities:

$$(x - y, y - z) \sim (x, -y, y, -z) \sim (x, -z) \sim (x - z);$$

and we infer from the strong associative law that

$$(x - y) + (y - z) = x - z.$$

Thus it follows from IX that the sums $(y - z) + (x - y)$ and $-z + x$ exist in A; and this implies the validity of the following vector similarities:

$$((y - z) + (x - y)) \sim (y, -z, x, -y) \sim (y, -z + x, -y).$$

Hence it follows from (S. 2) that the last of these vectors is reducible; and consequently at least one of the sums $y + (-z + x)$ and $(-z + x) - y$ exists in A; and it follows from IX and IX* that both these sums exist in A. Using the equation $-(-t) = t$ one deduces now successively from VII, IX, IX* the existence of the sums

$$(-z + x) + y, -z + (x + y), (x + y) - z, (x + y) + z.$$

Thus VIII is satisfied by the add A; and we have shown that (iv) is a consequence of (iii).

If Postulates I, II, VIII, X are satisfied by A, then we see as before that Postulates VI, VII are satisfied too; and it follows from §2, Theorem that A is an amalgam in the strict sense. This completes the proof.

4. Intrinsic characterization of the strong associative law. It is our objective in this section to characterize by inner properties the adds which satisfy the strong associative law. We precede this characterization by the proof of a couple of useful lemmas.

LEMMA 1. *Assume that Postulates* I *and* V *are satisfied by the add A.* (a) *If* (0) *is a reduction of the vector* $(v_1, \cdots, v_n)$, *and if* $2 < n$, *then there exists an integer i with the following properties*:

$0 < i < n - 1$, *the sum* $v_i + v_{i+1}$ *exists in A and* (0) *is a reduction of the contracted vector* $(v_1, \cdots, v_{i-1}, v_i + v_{i+1}, v_{i+2}, \cdots, v_n)$.

(b) (0) *is a reduction of the vector* $(v_1, \cdots, v_n)$ *with* $1 < n$ *if, and only if, there exists a reduction* (a) *of* $(v_1, \cdots, v_{n-1})$ *such that* $a + v_n = 0$.

Proof. If (0) is a reduction of the vector (a, b, c) of length 3, then either $(a+b)+c=0$ or $a+(b+c)=0$. It is a consequence of Postulate V that these two equations are equivalent conditions on the elements a, b, c. Consequently the sum $a+b$ exists in A; and (0) is a reduction of $(a+b, c)$. Thus Proposition (a) has been verified for $n=3$. Assume now that $3<n$ and that Proposition (a) has been verified for $n-1$. If (0) is a reduction of the vector $(v_1, \cdots, v_n)$, then there exists certainly an integer k with the following properties.

$0<k<n$, the sum v_k+v_{k+1} exists in A, and (0) is a reduction of the contracted vector $w=(\cdots, v_{k-1}, v_k+v_{k+1}, v_{k+2}, \cdots)$.

If k happens to be less than $n-1$, then k is the desired integer; and if $k=n-1$, then (0) is a reduction of the vector w whose length $n-1$ is at least 3. Thus we may apply the inductive hypothesis on

$$w=(v_1, \cdots, v_{n-2}, v_{n-1}+v_n).$$

Hence there exists an integer i with the following properties:

$0<i<n-2$, the sum v_i+v_{i+1} exists in A, and (0) is a reduction of the [twice contracted] vector

$$t=(v_1, \cdots, v_{i-1}, v_i+v_{i+1}, v_{i+2}, \cdots, v_{n-2}, v_{n-1}+v_n).$$

But then (0) is a reduction of the vector

$$s=(v_1, \cdots, v_{i-1}, v_i+v_{i+1}, v_{i+2}, \cdots, v_n),$$

of which t is a [simple] contraction. Thus we have found a suitable integer i in either case. This completes the induction and the proof of (a). (b) is readily deduced from (a) by complete induction.

Remark. By interchanging right and left in Lemma 1 another pair of correct assertions is obtained.

LEMMA 2. *If Postulates* I, IV *and* V *are satisfied by the add A, and if (0) is a reduction of the vector $(v_1, \cdots, v_n)$ with $1<n$, then (0) is also a reduction of the vector $(v_i, \cdots, v_n, v_1, \cdots, v_{i-1})$.*

Proof. Clearly it suffices to prove that (0) is a reduction of $(v_n, v_1, \cdots, v_{n-1})$ whenever (0) is a reduction of $(v_1, \cdots, v_n)$. But if (0) is a reduction of $(v_1, \cdots, v_n)$, then there exists, by Lemma 1, (b), a reduction (w) of $(v_1, \cdots, v_{n-1})$ such that $w+v_n=0$. It follows from Postulate IV that $0=v_n+w$. Hence (0) is a reduction of (v_n, w) and this vector is clearly a reduction of $(v_n, v_1, \cdots, v_{n-1})$ so that (0) is also a reduction of this latter vector, as we desired to prove.

THEOREM. *If Postulates* I *and* II *are satisfied by the add A, then the following properties of A are equivalent.*

(i) *The strong associative law.*

(ii) *Postulate* VI *and Property* (S. 2).

(iii) *Postulates* VI, VII *together with the following property:*

(S) *If the sums* $x_1 - x_2, \cdots, x_{n-1} - x_n$, $x_n - x_1$ *exist in* A, $3 < n$, *then the sum* $x_{i-1} - x_{i+1}$ *exists in* A *for some* i *such that* $1 < i < n$.

(iv) *Postulates* III, IV, V *together with the following property:*

(S. 0) *If* (0) *is a reduction of the vector* $(v_1, \cdots, v_n)$ *with* $1 < n$, *and if the difference* $v_{n-1} + v_n$ *exists in* A, *then* (0) *is also a reduction of the contracted vector* $(v_1, \cdots, v_{n-2}, v_{n-1} + v_n)$.

We note that the Postulates III, IV, V are consequences of the Postulates I, II, VI, VII [see Section **1**] and that it is therefore possible to use in Condition (iii. S) the symbol $-x$.

Proof. It is an immediate consequence of §**3**, Lemma 1 that (i) implies (ii). Assume next the validity of (ii). Suppose that the sums $a + b$, $b + c$ exist in A, and that at least one of the sums $(a + b) + c$ and $a + (b + c)$ exists in A. If, for instance, $a + (b + c)$ exists in A, then we have the vector similarities:

$$(a + (b + c)) \sim (a, b, c) \sim (a + b, c).$$

Since the first of these vectors has length 1 and the last one length 2, we infer from (S. 2) reducibility of the last of these vectors. Hence the sum $(a + b) + c$ exists in A; and this proves the validity of VII. Thus, I, II, VI, VII are satisfied by A; and it follows as usual that III, IV, V hold too. Hence we may use the symbol $-t$. Suppose now that the sums $x_1 - x_2, \cdots$, $x_{n-1} - x_n$, $x_n - x_1$ exist in A. Then we infer from §**1**, (2) the existence of $-(x_n - x_1) = -(-x_1) - x_n = x_1 - x_n$; and consequently we have the following vector similarities:

$$(x_1 - x_2, \cdots, x_{n-1} - x_n) \sim (x_1, -x_2, x_2, -x_3, \cdots, x_{n-1}, -x_n)$$
$$\sim (x_1, -x_n) \sim (x_1 - x_n).$$

The first of these vectors has length $n - 1$ and the last one length 1. Thus we infer from $2 < n$ and (S. 2) the reducibility of the first of these vectors. Consequently there exists an integer i such that $0 < i < n - 1$ and such that the sum $(x_i - x_{i+1}) + (x_{i+1} - x_{i+2})$ exists in A. This implies the validity of the following vector similarities:

$$((x_i - x_{i+1}) + (x_{i+1} - x_{i+2})) \sim (x_i, -x_{i+1}, x_{i+1}, -x_{i+2})$$
$$\sim (x_i, -x_{i+2}).$$

But then (S. 2) implies the reducibility of the last of these vectors; and this shows the existence of the sum $x_i - x_{i+2}$. Thus we have verified (S); and this shows that (iii) is a consequence of (ii).

Assume now the validity of (iii). Then we see as usual the validity of all the Posulates I to VII. Property (S. 0) is certainly satisfied by vectors of length 2; and it is equivalent with V for vectors of length 3; and thus we may make the inductive hypothesis that (S. 0) is satisfied by all vectors whose length is less than n and that $3 < n$. Suppose now that (0) is a reduction of the vector $(v_1, \cdots, v_n)$ and that the sum $v_{n-1} + v_n$ exists in A. We distinguish two possibilities.

Case 1. There exists an integer i with the following properties:

$0 < i < n - 2$, the sum $v_i + v_{i+1}$ exists in A, and (0) is a reduction of the contracted vector $(v_1, \cdots, v_{i-1}, v_1 + v_{i+1}, v_{i+2}, \cdots, v_n)$.

Since this contracted vector has length $n - 1$, we may apply the inductive hypothesis. Since $i + 1 < n - 1$, this implies that (0) is a reduction of the twice contracted vector

$$(v_1, \cdots, v_{i-1}, v_i + v_{i+1}, v_{i+2}, \cdots, v_{n-2}, v_{n-1} + v_n).$$

But then it is clear that (0) is likewise a reduction of the [once contracted] vector $(v_1, \cdots, v_{n-2}, v_{n-1} + v_n)$, as we claimed.

Case 2. If $0 < i < n - 2$, and if the sum $v_i + v_{i-1}$ exists in A, then (0) is not a reduction of the contracted vector

$$(v_1, \cdots, v_{i-1}, v_i + v_{i+1}, v_{i+2}, \cdots, v_n).$$

[Incidentally this can only happen for $n > 4$.]

In this case we deduce from Lemma 1, (a) inductively the existence of elements $w_{n-1}, \cdots, w_1$ with the following properties:

(a) the vector (w_i) is a reduction of the vector $(v_i, v_{i+1}, \cdots, v_{n-1})$;

(b) $w_{i-1} = v_{i-1} + w_i$;

(c) (0) is a reduction of the vector $(v_1, \cdots, v_{i-1}, w_i, v_n)$.

We note first two immediate consequences:

(a′) $w_{n-1} = v_{n-1}$;

(a″) $(v_1, \cdots, v_{i-1}, w_i, v_n)$ is a reduction of the vector $(v_1, \cdots, v_n)$.

From §1, (1) and from (b) we deduce $v_{i-1} = (v_{i-1} + w_i) - w_i = w_{i-1} - w_i$; and thus we have by §1, (2):

(b′) the sum $w_i - w_{i-1}$ exists in A for $1 < i < n$.

From (c) we deduce that (0) is a reduction of (w_1, v_n). Hence $0 = w_1 + v_n$ or

(c′) $w_1 = - v_n$.

Finally we deduce from the existence of $v_{n-1} + v_n$ the existence of $-(v_{n-1} + v_n) = - v_n - v_{n-1} = w_1 - w_{n-1}$ by §1, (2) and by (c′), (a′). Thus we have shown that the following sums exist in A:

$$w_1 - w_{n-1}, w_{n-1} - w_{n-2}, \cdots, w_2 - w_1.$$

It follows now from Condition (S) that either the sum $w_1 - w_{n-2}$ or one of the sums $w_{i+1} - w_{i-1}$ with $1 < i < n - 1$ exists in A. [This is trivial, if $n = 4$.]

We show next that the second possibility cannot arise. For otherwise some sum $w_{i+1} - w_{i-1}$ with $1 < i < n - 1$ would exist in A; and we would infer from §1, (2) that $w_{i-1} - w_{i+1}$ exists in A. It follows from (b) that

$$w_{i-1} - w_{i+1} = (v_{i-1} + w_i) - w_{i+1}.$$

But it follows from (b) [and §1, (1)] that $v_i = w_i - w_{i+1}$; and now we deduce from Postulates VI, VII that

$$\begin{aligned} w_{i-1} &= (w_{i-1} - w_{i+1}) + w_{i+1} = [(v_{i-1} + w_i) - w_{i+1}] + w_{i+1} \\ &= [v_{i-1} + (w_i - w_{i+1})] + w_{i+1} = (v_{i-1} + v_i) + w_{i+1}. \end{aligned}$$

Now we deduce from (c) that (0) is a reduction of the vector

$$(v_1, \cdots, v_{i-2}, w_{i-1}, v_n) = (v_1, \cdots, v_{i-2}, (v_{i-1} + v_i) + w_{i+1}, v_n);$$

and this vector is a contraction of the vector

$$(v_1, \cdots, v_{i-2}, v_{i-1} + v_i, w_{i+1}, v_n).$$

But it follows from (a) that the last vector is a reduction of

$$(v_1, \cdots, v_{i-2}, v_{i-1} + v_i, v_{i+1}, \cdots, v_{n-1}, v_n),$$

so that (0) has been shown to be a reduction of this vector contrary to the hypothesis of Case 2.

Thus we have shown the existence of the sum $w_1 - w_{n-2}$. Consequently we may deduce from (c′) the existence of the sum

$$-(w_1 - w_{n-2}) = w_{n-2} - w_1 = w_{n-2} + v_n.$$

Since the sum $v_{n-1} + v_n$ exists in A, we may now deduce from (b), (a′) and the Postulates VI, VII that

$$w_{n-2} + v_n = (v_{n-2} + w_{n-1}) + v_n = (v_{n-2} + v_{n-1}) + v_n$$
$$= v_{n-2} + (v_{n-1} + v_n).$$

It follows from (c) that (0) is a reduction of the vector

$$(v_1, \cdots, v_{n-3}, w_{n-2}, v_n)$$

whose length is $n - 1$. Since the sum $w_{n-2} + v_n$ exists in A, it follows from the inductive hypothesis that (0) is likewise a reduction of the vector

$$(v_1, \cdots, v_{n-3}, w_{n-2} + v_n) = (v_1, \cdots, v_{n-3}, v_{n-2} + (v_{n-1} + v_n)).$$

But this vector is a contraction of $(v_1, \cdots, v_{n-3}, v_{n-2}, v_{n-1} + v_n)$ so that (0) is a reduction of this last vector, as we desired to show. This completes the inductive proof of (S. 0); and thus we have shown that (iv) is a consequence of (iii).

Assume now the validity of (iv). Then we prove first:

(iv. 1) If (0) is a reduction of the vector $v = (v_1, \cdots, v_n)$, then (0) is a reduction of every contraction of v.

If the sum $v_i + v_{i+1}$ exists in A, $0 < i < n$, then we deduce first from Lemma 2 that (0) is a reduction of the vector

$$v' = (v_{i+2}, \cdots, v_n, v_1, \cdots, v_i, v_{i+1});$$

and we deduce from Condition (S. 0) that (0) is a reduction of the contraction

$$v'' = (v_{i+2}, \cdots, v_n, v_1, \cdots, v_{i-1}, v_i + v_{i+1});$$

and consequently it follows from Lemma 2 again that (0) is a reduction of the contraction $(v_1, \cdots, v_{i-1}, v_i + v_{i+1}, x_{i+2}, \cdots, v_n)$ of v, as we desired to show.

(iv. 2) Suppose that $v \sim w$. Then (0) is a reduction of v if, and only if, (0) is a reduction of w.

To prove this we note first the following immediate consequence of (iv. 1): If the vector r is a contraction of the vector s, then (0) is a reduction of r if, and only if, (0) is a reduction of s; and from this fact the proposition (iv. 2) follows by complete induction.

Suppose now that $(a_1, \cdots, a_n)$ is an irreducible vector over A of positive length n; and that a is an element in A such that

$$(a_1, \cdots, a_n) \sim (a).$$

Since the Postulates I to V are satisfied by A, we may make use of the operation $-t$. Consequently we find that

$$(0) \sim (a, -a) \sim (a_1, \cdots, a_n, -a) = v.$$

It follows from (iv. 2) that (0) is a reduction of the vector v; and now it follows from Lemma 1, (b) and the irreducibility of the vector $(a_1, \cdots, a_n)$ that $n = 1$ and $a_1 - a = 0$. But this implies [because of Postulates I to V] that $a_1 = a$. Thus we have shown the validity of the strong associative law; and this completes the proof.

COROLLARY 1. *The add A is an amalgam satisfying the strong associative law if, and only if, A satisfies Postulates* I, II, VI, VII *and*

(S*) *the existence of the sums $x_1 + x_2, \cdots, x_{n-1} + x_n, x_n + x_1$ with $3 < n$ implies the existence of a sum $x_{i-1} + x_{i+1}$ with $1 < i < n$.*

Proof. If A is an amalgam satisfying the strong associative law, then A satisfies Postulates I, II, VI, VII [§3, Lemma 1]; and the existence of $a + b$ implies clearly the existence of $a - b$. Thus the existence of the sums $x_1 + x_2, \cdots, x_{n-1} + x_n, x_n + x_1$ implies the existence of the differences $x_1 - x_2, \cdots, x_{n-1} - x_n, x_n - x_1$; and it follows from the preceding Theorem that at least one of the differences $x_{i-1} - x_{i+1}$ exists in A. But this implies the existence of the sum $x_{i-1} + x_{i+1}$ in the amalgam A. Thus we have shown the necessity of our conditions.

Assume conversely the validity of Postulates I, II, VI, VII and (S*). If $a + b$ exists in A, then we are assured of the existence of the following sums: $a + b$, $b - b\ [= 0]$, $-b - a\ [= -(a + b)]$, $-a + a$; and it follows from (S*) that at least one of the sums $a - b$ and $b - a$ exists in A. But the existence of $b - a$ implies the existence of $-(b - a) = a - b$; and thus we have shown that the existence of $a + b$ implies the existence of $a - b$. This Property IX* implies together with (S*) the property (S) of the preceding Theorem, so that the strong associative law holds in A. One deduces furthermore from IX* as usual the Postulates IX and X; and now it follows from §3, Theorem that A is an amalgam in the strict sense satisfying the strong associative law.

COROLLARY 2. *The amalgam A [in the weak sense] satisfies the strong associative law, if*

(a) *to any three distinct maximal subgroups X, Y, Z of A there exists a subgroup S of A such that $X \cap Y$, $Y \cap Z$, $Z \cap X$ are all part of S; and*

(b) *to any $n > 3$ maximal subgroups $X_1, \cdots, X_{n-1}, X_n = X_0$ of A there exists a subgroup T of A and an integer i with $0 < i < n - 1$ such that $X_{i-1} \cap X_i$ and $X_{i+1} \cap X_{i+2}$ are both part of T.*

Proof. It is a consequence of (a) and §3, Lemma 2 that A is an amalgam in the strict sense; and we deduce from (b) that Corollary 1, (S*) is satisfied by A. Thus the validity of the strong associative law may be inferred from Corollary 1.

Remark 1. It is an open question whether Condition (b) is necessary, though certain parts of (b) may be shown to be necessary. It is easy, however, to construct examples which show that it does not suffice to require the validity of (b) for bounded n only.

Remark 2. Most of the known criteria for the validity of the strong associative law are easily seen to be special cases of Corollary 2; see in particular Baer (1) §8, Theorem 1 and Corollary 1.

5. Generalization of H. Neumann's reduction theorem. We suppose throughout this section that the add A satisfies Postulates I to V. Then A possesses a uniquely determined null element, the operation $-x$ is well defined, and cross cuts of subgroups are subgroups of A.

If M is a maximal subgroup of A, then we form the subgroup $C(M)$ of M which is generated by all the subgroups $M \cap X$, where X ranges over all the maximal subgroups, not M, of A. The join $D(A)$ of all these subgroups $C(M)$ is a subadd of A with a number of important properties:

$$C(M) = M \cap D(A) \text{ for every maximal subgroup } M \text{ of } A.$$

Consequently $D(A)$ contains $-x$, whenever it contains x; and it contains $x + y$, whenever $x + y$ exists in A and x and y belong to $D(A)$. Hence $D(A)$ is an amalgam [in the weak sense]; and it is an amalgam in the strict sense whenever A has this property [see also Baer (1), §8, Lemma 3]. It can also be shown that the strong associative law is satisfied by the amalgam A if, and only if, it is satisfied by $D(A)$ [Baer (1), §8, Theorem 2; Neumann (1), 5. 0. Theorem p. 599]; and it is our objective to prove an analogous property for the iterates of $D(A)$ which includes the result just mentioned as a special case.

We define the iterates of the operation D in the following obvious fashion:

$A = D_0(A)$,

$D_{\nu+1}(A) = D[D_\nu(A)]$,

$D_\sigma(A($ = intersection of all the $D_\nu(A)$ with $\nu < \sigma$ in case σ is a limit ordinal.

One proves easily by complete [transfinite] induction that the $D_\nu(A)$ form a descending chain of subadds of A with the closure property:

$D_\nu(A)$ contains $-x$ whenever it contains x; and it contains $x + y$ whenever it contains x and y and the sum $x + y$ exists in A.

The $D_\nu(A)$ with $0 < \nu$ are all amalgams of their subgroups; and if A is an amalgam in the strict sense, then all the $D_\nu(A)$ are amalgams in the strict sense.

Finally we note the existence of an ordinal τ such that $D_\tau(A) = D_{\tau+1}(A) = D[D_\tau(A)]$. It is of interest to note that there exist [easily constructed] amalgams A in the strict sense where this ordinal τ cannot be finite.

THEOREM. *If A is an amalgam in the strict sense, then the following properties of A are equivalent.*

(i) *The strong associative law is satisfied by A.*

(ii) *The strong associative law is satisfied by every $D_\nu(A)$.*

(iii) *The strong associative law is satisfied by at least one $D_\nu(A)$.*

Proof. We note first that every $D_\nu(A)$ is an amalgam in the strict sense. Assume now the validity of the strong associative law in A; and consider elements $a_1, \cdots, a_n$ in $D_\rho(A)$ with the property: the sums $a_1 + a_2, \cdots, a_{n-1} + a_n$, $a_n + a_1$ exist in $D_\rho(A)$ and $3 < n$. Since the elements a_i and their sums are in A, and since the strong associative law holds in A, we may deduce from §4, Corollary 1, (S*) the existence of one of the sums $a_{i-1} + a_{i+1}$ with $1 < i < n$ in A. But $D_\rho(A)$ is closed in A with respect to addition; and so this sum exists in $D_\rho(A)$. Consequently Condition (S*) is satisfied by $D_\rho(A)$; and we deduce from §4, Corollary 1 the validity of the strong associative law in $D_\rho(A)$. Thus we have shown that (ii) is a consequence of (i); and (iii) is just a special case of (ii).

Assume now the validity of (iii). Then we want to show the validity of Condition (S*) of §4, Corollary 1 in A. If this were not true, then there would exist a minimal integer $m > 3$ with the following properties:

(T. m) there exist elements $w_1, \cdots, w_m$ in A such that the sums $w_1 + w_2, \cdots, w_{m-1} + w_m, w_m + w_1$ exist in A, whereas none of the sums $w_{i-1} + w_{i+1}$ with $1 < i < m$ exists in A.

We show next that these elements w_i, satisfying (T. m), belong to every

$D_\nu(A)$. This is certainly true for $\nu = 0$; and thus we may make the inductive hypothesis:

$$0 < \nu \text{ and the } w_i \text{ belong to every } D_\sigma(A) \text{ with } \sigma < \nu.$$

If ν happens to be a limit ordinal, then $D_\nu(A)$ is the intersection of all the $D_\sigma(A)$ with $\sigma < \nu$ so that the w_i belong to $D_\nu(A)$ too [because of the inductive hypothesis]. If $\nu = \rho + 1$ is not a limit-ordinal, then it follows from the inductive hypothesis that the w_i belong to $D_\rho(A)$. Since $D_\rho(A)$ is closed in A with respect to addition, it follows from (T. m) that all the sums $w_1 + w_2, \cdots, w_{m-1} + w_m, w_m + w_1$ belong to $D_\rho(A)$. Since $D_\rho(A)$ is an amalgam of its subgroups, there exist maximal subgroups S_i of $D_\rho(A)$ such that w_i, w_{i+1} are in S_i for $0 < i < m$ and such that w_m, w_1 belong to S_m. If $S_i = S_{i+1}$ for some positive $i < m - 1$, then w_i, w_{i+1}, w_{i+2} would be in the same subgroup of A so that the sum $w_i + w_{i+2}$ would exist contrary to (T. m). Hence

$$S_i \neq S_{i+1} \text{ for } 0 < i < m - 1.$$

Next we show that $S_{m-1} \neq S_m \neq S_1$. To do this we distinguish two cases.

Case 1. $m = 4$. If $S_3 = S_4$, then w_3, w_4, w_1 would be in the same subgroup of A so that the sum $w_1 + w_3$ would exist in A, contrary to (T. 4); and if $S_4 = S_1$, then w_4, w_1, w_2 would belong to the same subgroup of A so that the sum $w_2 + w_4$ would exist in A, contrary to (T. 4). Hence $S_3 \neq S_4 \neq S_1$, as we claimed.

Case 2. $4 < m$. If $S_{m-1} = S_m$, then w_{m-1}, w_m, w_1 belong to the same subgroup of A so that the sum $w_{m-1} + w_1$ exists in A; and the elements $w_1, \cdots, w_{m-1}$ satisfy (T. m — 1). If $S_m = S_1$, then w_m, w_1, w_2 belong to some subgroup of A so that the sum $w_m + w_2$ exists in A; and the elements $w_2, \cdots, w_m$ satisfy (T. m — 1). Either way we are led to a contradiction with our minimal choice of m. Hence $S_{m-1} \neq S_m \neq S_1$, as we claimed.

Since the S_i are maximal subgroups of the amalgam $D_\rho(A)$, the operation $C(S_i)$ is well defined with respect to $D_\rho(A)$. The element w_1 belongs to the cross cut $S_1 \cap S_m$ of two distinct maximal subgroups of $D_\rho(A)$. Hence w_1 is in $C(S_1)$ and therefore in $D[D_\rho(A)]$. Likewise w_i with $1 < i \leq m$ belongs to the cross cut $S_{i-1} \cap S_i$ of two distinct maximal subgroups of $D_\rho(A)$ and hence to $D[D_\rho(A)]$. Thus we have shown that every w_i belongs to $D[D_\rho(A)] = D_{\rho+1}(A) = D_\nu(A)$; and this completes our inductive proof.

We have shown now that Condition (S*) of §4, Corollary 1 is not satisfied by any $D_\nu(A)$, if it is not satisfied by A. But it follows from (iii) and §4, Corollary 1 that this Condition (S*) is satisfied by at least one

$D_\nu(A)$. Hence it is satisfied by A too; and this implies the validity of the strong associative law in A. Hence (i) is a consequence of (iii); and this completes the proof.

Remark 1. If any two distinct maximal subgroups of the amalgam A meet in the same subgroup U, then $D(A) = U$; and it follows from the preceding theorem that the strong associative law is satisfied by A. Thus one sees that Schreier's Theorem is a special case of the preceding theorem. On the other hand it does not seem likely that a criterion like §4, Corollary 2 could be deduced from our result.

Remark 2. If A is an amalgam in the strict sense, then there exists some ordinal τ such that the amalgam $D_\tau(A) = B$ is an amalgam in the strict sense which satisfies $B = D(B)$. It follows from our theorem that the strong associative law holds in A if, and only if, it holds in B. Thus it suffices to find criteria for the validity of the strong associative law in amalgams B [in the strict sense] which satisfy in addition the condition $B = D(B)$.

University of Illinois,
Urbana, Illinois.

BIBLIOGRAPHY.

Reinhold Baer.

(1) "Free sums of groups and their generalizations. An analysis of the associative law," *American Journal of Mathematics*, vol. 71 (1949), pp. 706-742.

B. H. Neuman and Hanna Neumann.

(1) "A remark on generalized free products," *Journal of the London Mathematical Society*, vol. 25 (1950), pp. 202-204.

Hanna Neumann.

(1) "Generalized free products with amalgamated subgroups," *American Journal of Mathematics*, vol. 70 (1948), pp. 590-625.

FREE SUMS OF GROUPS AND THEIR GENERALIZATIONS III.*

By Reinhold Baer.

In the first communication of this series we have shown that the appropriate framework for an investigation of free sums of groups with amalgamated subgroups is provided by the concept of add [= system with an incomplete addition]. In particular we have stressed the central importance of what we called the strong associative law. This may be restated in a form appropriate for the purposes of this introduction as follows: similar irreducible vectors are equal, if one of them has length 1. The connection of such a requirement with the so-called word [or identity] problem is obvious, since it requires a trivial solution in a special case. It is our objective in the present investigation to characterize adds meeting the following still stronger requirements: (a) the class of adds where similar irreducible vectors are equal [Section **2**]; (b) the class of adds where similar irreducible vectors have equal length and where similar vectors of length 1 are equal [Section **1**]. In adds of class (a) the word problem has a trivial solution, and in adds of class (b) a simple solution of the word problem exists [Section **3**]. In Section **4** we establish finally a striking and surprising relation between the word problem and the transformation problem.

Results and concepts of our second communication are used freely in the present investigation. References to this paper are indicated by the prefix II [like II, §4, Lemma 1]. The postulates I to X which we enumerated in II, §**1** are used so often that we refer to them just by their roman numeral without any further reference.

1. Invariance of length of irreducible reductions. The principal objective of this section is the proof of the equivalence of three properties of adds. Before enunciating this theorem it will be convenient to discuss some of the concepts and properties occurring in this theorem.

The Associative Law (T) consists of the Property (S. 1) of the strong associative law [or associative law (S) as we shall term it in the future] and of the following invariance property:

---

* Received June 6, 1949.

(L) *Similar irreducible vectors have the same length.*

It is clear that the Property (S. 2) is a special case of (L), since it requires the validity of (L), if at least one of the vectors has length 1. Thus the associative law (S) is a consequence of the associative law (T).

The property (L) is easily seen to be equivalent to the apparently weaker property

(L′) *If v' and v'' are irreducible reductions of the vector v, then v' and v'' have the same length.*

Proof of the equivalence of Properties (L) *and* (L′): if v' and v'' are reductions of v, then they are similar vectors; and thus (L′) is a consequence of (L). Suppose conversely the validity of (L′) and the similarity of the irreducible vectors u and w. Then there exist vectors $v(i)$ such that $u = v(0)$, $w = v(2n)$ and $v(2i)$, $v(2i+2)$ are reductions of $v(2i+1)$ for $0 \leq i < n$. [Note that some of these vectors may be equal.] Since u and w are irreducible vectors, and since every vector possesses an irreducible reduction, we may assume without loss in generality that the vectors $v(2j)$ are irreducible. But then it follows from (L′) that $v(2i)$ and $v(2i+2)$ for $0 \leq i < n$ have the same length; and thus u and w have the same length. Hence (L) is a consequence of (L′).

We note finally that Property (S. 1) is not a consequence of Property (L), as may be seen from the example of any system with a unique, but non-associative, addition where all irreducible vectors have length 1. That (S. 1) does not imply (S. 2) nor (L) is likewise easily seen.

THE POSTULATE XI consists of three parts:

(a) *If the sums $a + b$, $b + c$, $c + d$ exist in A, then at least one of the sums $a + (b + c)$ and $(b + c) + d$ exists in A.*

(b) *If the sums $b + c$, $c + d$ and $a + (b + c)$ exist in A, then at least one of the sums $a + b$ and $(b + c) + d$ exists in A.*

(c) *If the sums $a + b$, $b + c$ and $(b + c) + d$ exist in A, then at least one of the sums $a + (b + c)$ and $c + d$ exists in A.*

In certain instances it is possible to deduce the properties (b), (c) from (a); but whether or not this is true in general, the author does not know.

CONNECTED CONTRACTIONS. This concept is not of as great an intrinsic interest as the preceding ones; but it will prove very helpful in several proofs. We define it as follows:

If $v(1), \cdots, v(m)$ *are contractions of the vector* v*, and if* $v(i)$ *and* $v(i+1)$ *possess, for* $0 < i < m$*, a common contraction, then* $v(1)$ *and* $v(m)$ *are connected contractions of* v*.*

We shall utilize this concept for $m = 2, 3$ only. The reason for this is contained in the following useful proposition.

LEMMA. *If there exist contractions of the vector* v *which are not connected, then* v *satisfies one of the following two (exclusive) properties:*

(a) $v = v' + (a, b, c) + v''$ *where* $v' + (a)$ *and* $(c) + v''$ *are irreducible vectors and where the sums* $a + b$, $b + c$ *exist in* A*, though the existence of the sums* $a + (b + c)$ *and* $(a + b) + c$ *would imply* $a + (b + c) \neq (a + b) + c$.

(b) $v = w' + (p, q, r, s) + w''$ *where* $w' + (p)$ *and* $(s) + w''$ *are irreducible vectors and where the sums* $p + q$, $q + r$, $r + s$ *exist in* A*, though the existence of the sums* $p + (q + r)$ *and* $(p + q) + r$ *would imply* $p + (q + r) \neq (p + q) + r$ *and the existence of the sums* $q + (r + s)$ *and* $(q + r) + s$ *would imply* $q + (r + s) \neq (q + r) + s$.

Proof. Our arguments are based on the following obvious constructions:

(A) If $v = x + (a, b) + y + (c, d) + z$, and if the sums $a + b$ and $c + d$ exist in A, then $x + (a + b) + y + (c + d) + z$ is a common contraction of the contractions $x + (a + b) + y + (c, d) + z$ and $x + (a, b) + y + (c + d) + z$ of the vector v.

(B) If $v = x + (a, b, c) + y$, if the sums $a + b$, $b + c$, $a + (b + c)$, $(a + b) + c$ exist in A, and if $a + (b + c) = (a + b) + c = s$, then $x + (s) + y$ is a common contraction of the contractions $x + (a + b, c) + y$ and $x + (a, b + c) + y$ of the vector v.

If the vector $v = (v_1, \cdots, v_n)$ possesses contractions which are not connected, then v possesses at least two contractions. Consequently there exists an integer k such that $1 < k < n$ and integers $f(1), \cdots, f(k)$ such that $0 < f(1) < \cdots < f(j) < f(j + 1) < \cdots < f(k) < n$ with the following properties:

the sum $v_i + v_{i+1}$ exists in A if, and only if, $i = f(j)$ for some j with $1 \leq j \leq k$.

If we denote now by $v(j)$ the vector which arises from v by substituting for $v_{f(j)}$, $v_{f(j)+1}$ their sum, then the vectors $v(1), \cdots, v(k)$ constitute all the contractions of v.

It follows from (A) that $v(i)$ and $v(j)$ have a common contraction whenever $f(i)+1<f(j)$. This implies in particular that $v(1)$ and $v(i)$ possess a common contraction for $2<i$ and that $v(j)$ and $v(k)$ possess a common contraction for $j<k-1$. Consequently any two contractions of v are connected if $3<k$; or if $k=3$, but $f(1)+2<f(3)$. The existence of unconnected contractions of v implies therefore that $k=2$ or 3, $f(2)=f(1)+1$ and, in case $k=3$, then $f(3)=f(2)+1=f(1)+2$. The validity of our lemma is now readily deduced from (B).

THEOREM 1. *The following properties of the add A are equivalent.*

(i) *The associative law* (T).

(ii) *The postulates* VI, VII, XI.

(iii) *If the contractions v' and v'' of the vector v are not both irreducible, then they are connected contractions of v.*

Proof. Assume first the validity of the associative law (T). This implies the associative law (S) of which VI, VII are consequences [II, §3, Lemma 1]. If the sums $a+b$, $b+c$, $c+d$ exist in A, then we have the vector similarities:

$$(a+b, c+d) \sim (a, b, c, d) \sim (a, b+c, d).$$

Hence it follows from the invariance rule (L) that the last of these vectors is reducible; and this proves the existence of at least one of the sums $a+(b+c)$ and $(b+c)+d$; i.e. the validity of XI, (a). If next the sums $a+b, b+c$, $(b+c)+d$ exist in A, then we obtain the vector similarities:

$$(a, (b+c)+d) \sim (a, b, c, d) \sim (a+b, c, d).$$

As before we deduce from (L) the reducibility of the last of these vectors and therefore the existence of at least one of the sums $(a+b)+c$ and $c+d$ in A so that the validity of XI, (c) has been established; and XI, (b) is verified similarly. Thus we have shown that (ii) is a consequence of (i).

Assume now the validity of (ii). Then we prove (iii) in the following slightly stricter form:

(iii*) *If the vector v possesses contractions which are not connected, then all its contractions are irreducible.*

If the vector v possesses contractions which are not connected, then it follows from Lemma 1 that v possess either two or three contractions. If v

possesses just two contractions, then we infer from Lemma 1, (a) and Postulates VI, VII, that v has the following form:

$v = v' + (a, b, c) + v''$ where $v' + (a)$ and $(c) + v''$ are irreducible vectors and where the sums $a + b$, $b + c$ exist in A, though none of the sums $a + (b + c)$, $(a + b) + c$ exists in A.

If $v' + (a + b, c) + v''$ were reducible, then we would have $v' = w + (t)$ where $t + (a + \mathrm{b})$ would exist in A. But from the existence of the sums $a + b$, $b + c$ and $t + (a + b)$ in A we infer by XI, (b) the existence of at least one of the sums $t + a$ and $(a + b) + c$, which is impossible. Thus the contraction $v' + (a + b, c) + v''$ of v is irreducible; and the irreducibility of the other contraction $v' + (a, b + c) + v''$ of v is derived similarly from XI, (c).

If v possesses just three contractions, then we infer from Lemma 1, (b) and Postulates VI, VII that v has the following form:

$v = w' + (p, q, r, s) + w''$ where $w' + (p)$ and $(s) + w''$ are irreducible vectors and where the sums $p + q$, $q + r$, $r + s$ exist in A, though none of the sums $p + (q + r)$, $(p + q) + r$, $(q + r) + s$, $q + (r + s)$ exists in A.

This clearly contradicts XI, (a).

Thus we have verified the validity of (iii*) and shown that (iii) is a consequence of (ii).

Assume finally the validity of (iii). We first verify (L), and to do this it suffices to prove (L′). This we do by complete induction with respect to the length n of the vector v. It is trivially true, if v has length 1; and thus we may assume that $1 < n$ and that (L′) is true for all vectors whose length is less than n. If the vector v of length n is irreducible, then our contention is again trivially true. Thus we assume finally that v is reducible and that v' and v'' are irreducible reductions of v. Consequently there exist contractions w' and w'' such that v' is a reduction of w' and v'' is a reduction of w''. If w' and w'' are both irreducible, then $v' = w'$ and $v'' = w''$ are both of length $n - 1$, as we wanted to show. If w' and w'' are not both irreducible, then they are, by (iii), connected; and there exist consequently contractions $w(0), \cdots, w(k)$ of v such that $w' = w(0)$, $w'' = w(k)$ and such that $w(i)$ and $w(i + 1)$ possess for $0 \leq i < k$ a common contraction $u(i)$. Since $w(i)$ has length $n - 1$, it follows from the inductive hypothesis that all the irreducible reductions of $w(i)$ have the same length $m(i)$. But $u(i)$ has an irreducuible reduction which is necessarily a reduction of both $w(i)$ and of $w(i + 1)$. Consequently $m(i) = m(i + 1)$ for every i, so that in particular

$m(0) = m(k)$. But $m(0)$ is the length of v', and $m(k)$ is the length of v''. Hence v' and v'' have the same length, as we wanted to show. This completes the inductive proof of (L$'$). Hence (L) too is satisfied in A.

Next we prove:

(S. 1$'$) *If* (a) *is a reduction of the vector* v, *then* (a) *is the only irreducible reduction of* v.

This is trivially true, if v is of length 1. Thus we may assume that v be of length $n > 1$ and that we have verified (S. 1$'$) for vectors whose length is less than 1. It follows from (L) that all irreducible reductions of v have length n. Suppose therefore that (b) is some irreducible reduction of v. Then there exist contractions u and w of v such that (a) is a reduction of u and (b) is a reduction of v. If u and w are irreducible, then $v = (r, s)$ is of length 2 and $(a) = (r + s) = (b)$ so that $a = r + s = b$ too. If u and w are not irreducible, then we infer from (iii) the existence of contractions $w(i)$ of v such that $u = w(0)$, $w = w(k)$ and such that $w(i)$ and $w(i+1)$ possess a common contraction $u(i)$. Since the $w(i)$ are of length $n - 1$, and since all their irreducible reduction have length 1 [as reductions of v], it follows that $w(i)$ has one and only one irreducible reduction (a_i). But (a_i) is a reduction of $u(i)$, as is (a_{i+1}); and so it follows that $a_i = a_{i+1}$. This implies in particular the equality of $a = a_0$ and $a_k = b$, so that the inductive proof of (S. 1$'$) has been completed. From (S. 1$'$) one deduces readily (S. 1); and thus we have shown that (i) is a consequence of (iii). This completes the proof of our theorem.

COROLLARY. *If Postulates* I *and* II *are satisfied by the add* A, *then the associative law* (T) *is a necessary and sufficient condition for the validity of Postulates* VI, VII *and*

XI-. *If* $a - b$, $b - c$ *and* $c - d$ *exist in* A, *then* $a - c$ *or* $b - d$ *exists in* A.

Proof. If (T) is satisfied by A, then we deduce the validity of VI, VII from Theorem 1. If furthermore $a - b$, $b - c$, $c - d$ exist in A, then we have

$$(a - b, b - c, c - d) \sim (a, -b, b, -c, c, -d) \sim (a, -d).$$

Since the first of the vectors has length 3 and the last one has length 2, it follows from (T) [invariance of length] that at least one of the sums $(a - b) + (b - c)$ and $(b - c) + (c - d)$ exists in A. But $(a - b) + b = a$ by VI, VII; and thus it follows from VI, VII that $(a - b) + (b - c) = a - c$, if it exists; and similarly one sees that $(b - c) + (c - d) = b - d$, if it exists. Hence XI- is a consequence of (T).

Assume conversely the validity of VI, VII and XI-. Then we are assured of the validity of Postulates I to V [II, §1]. If the sums $a + b$, $b + c$, $c + d$ exist in A, then it follows from I to VII that the differences $(a + b) - b$, $b - (-c)$, $-c - [-(c + d)]$ exist in A; and it follows from XI- that $(a + b) - (-c) = (a + b) + c = a + (b + c)$ or $b - [-(c + d)] = b + (c + d) = (b + c) + d$ exists in A. Thus XI, (a) is true. If the sums $a + b$, $b + c$ and $(b + c) + d$ exist in A, then we infer from Postulates I to VII the existence of the differences $a - (-b)$, $-b - [-(b + c)]$, $-(b + c) - [-((b + c) + d)]$; and it follows from XI- that at least one of the differences $a - [-(b + c)] = a + (b + c)$ or $-b - [-((b + c) + d)] = -b + [(b + c) + d] = c + d$ [because of $-b + (b + c) = c$] exists in A. Thus XI, (c) is true too; and the validity of XI, (b) is verified likewise. Hence VI, VII, XI are true; and (T) is a consequence of Theorem 1.

THEOREM 2. *The following properties of the add* A *are equivalent.*

(i) *A is the amalgam of its subgroups and satisfies the associative law* (T).

(ii) *A is the amalgam of its subgroups and satisfies the following condition:*

(D) *If X, Y, Z are maximal subgroups of A, then $X \cap Y \leq Y \cap Z$ or $Z \cap Y \leq Y \cap X$.*

(iii) *A satisfies Postulates* I, II, VI, VII *and*

XI*. *If the sums $a + b$, $b + c$, $c + d$ exist in A, then at least one of the sums $a + c$ and $b + d$ exists in A.*

Proof. Assume first the validity of (i). Then A is an amalgam which satisfies the strong associative law; and A is consequently an amalgam in the strict sense [II, §3, Theorem]. Suppose now that X, Y, Z are maximal subgroups of A; and that $X \cap Y \not\leq Y \cap Z$ and $Z \cap Y \not\leq Y \cap X$. Then there exists an element a in $X \cap Y$, but not in $Y \cap Z$. Since a is not in the maximal subgroup Z of A, there exists [II, §2, Corollary 1, (c)] an element c in Z such that the sum $a + c$ does not exist in A. Likewise there exists an element b in $Z \cap Y$ and an element d in X such that the sum $b + d$ does not exist in A. Since d, a are in X, a, b in Y and b, c in Z, we are assured of the existence of the sums $d + a$, $a + b$, $b + c$. Then it follows from Postulate XI, (a) [which is true because of (T) and Theorem 1] that at least one of the sums $d + (a + b)$ and $(a + b) + c$ exists in A. If the first of these sums exists in A, then we are assured of the existence of the sums $a + d$, $d + (a + b)$, $(a + b) + a$ in the amalgam A in the strict sense; and hence it follows [II, §2, Definition 2, (A. 3)] that $a, d, a + b$ are contained in some subgroup of A. But this subgroup would contain d and b too, which is impossible. If the sum $(a + b) + c$ would exist in A, then we would show likewise that a and c belong to the same subgroup of A, which is impossible too. Thus we are led to a contradiction which proves the validity of Condition (D). Hence (ii) is a consequence of (i).

Assume now the validity of (ii). If X, Y, Z are maximal subgroups of A, then it follows from (D) that the cross cuts $X \cap Y$, $Y \cap Z$, $Z \cap X$ are all contained in one of the subgroups X, Y, Z. Hence A is an amalgam in the strict sense [II, §3, Lemma 2]. Consequently the Postulates I to X are satisfied by A [II, §2]. Assume now the existence of the sums $a + b$, $b + c$, $c + d$ in A. Then there exist maximal subgroups X, Y, Z of A such that a, b are in X, b, c in Y and c, d in Z. Hence b belongs to $X \cap Y$ and c to $Y \cap Z$. If $X \cap Y \leq Y \cap Z$, then b is in Z, as are c and d so that the sum $b + d$ exist in A. If $X \cap Y \not\leq Y \cap Z$, then it follows from (D) that $Z \cap Y \leq Y \cap X$. Hence c belongs to X, as do a and b, so that the sum $a + c$ exists in A. Thus we have verified the validity of XI*; and we have shown that (iii) is a consequence of (ii).

Assume finally the validity of (iii). Then we are assured of the validity of Postulates I to VII and may use the operation $-x$ as usual [II, §1]. We show first the validity of

IX*. *The existence of* $a + b$ *implies the existence of* $a - b$.

If $a + b$ exists in A, then we are assured of the existence of the sums $a + b$, $b - b$ $[= 0]$, $-b - a$ $[= -(a + b)]$ in A. This implies by XI* the existence of at least one of the sums $a - b$ and $b - a$. But $-(b - a) = -(-a) - b = a - b$, so that $a - b$ exists in either case.

If the sums $x_1 - x_2, \cdots, x_{n-1} - x_n$, $x_n - x_1$ exist in A, and if $3 < n$, then we infer from IX* the existence of the sums $x_1 + x_2$, $x_2 + x_3$, $x_3 + x_4$. It follows from XI* that at least one of the sums $x_1 + x_3$ and $x_2 + x_4$ exists in A; and hence it follows from IX* that $x_1 - x_3$ or $x_2 - x_4$ exists in A. Thus we have verified the validity of II, §4, Theorem, (iii) and this shows the validity of the strong associative law.

From the existence of $a - a$ $[= 0]$ and IX* we infer the existence of $a - (-a) = a + a$. Thus X is satisfied by A. If $a + b$ exists in A, then we deduce from IX* successively the existence of $a - b$, $-(a - b) = b - a$, $b - (-a) = b + a$, so that IX is valid in A. Now it is a consequence of

II, §3, Theorem that A is an amalgam in the strict sense; and this implies in particular the validity of the postulates I to X [II, §2].

If the sums $a+b$, $b+c$, $c+d$ exist in A, then we infer from XI* that at least one of the sums $a+c$ and $b+d$ exists in A. If $a+c$ exists in A, then we are assured of the existence of the sums $a+b$, $b+c$, $a+c$; and it follows from VIII that $(a+b)+c=a+(b+c)$ [by VI, VII] exists in A; and if $b+d$ exists in A, then we see likewise the existence of $(b+c)+d$. Thus XI, (a) holds in A.

If the sums $a+b$, $b+c$ and $(b+c)+d$ exist in A, then we infer from the properties of amalgams the existence of the sums

$$a+b,\ b+(b+c),\ (b+c)+d.$$

Hence it follows from XI* that at least one of the sums $a+(b+c)$ and $b+d$ exists in A. In the latter case we are assured of the existence of the sums $d+b$, $b+(b+c)$, $(b+c)+d$; and it follows from the properties of amalgams in the strict sense [(A. 3)] that $d, b, b+c$ belong to some subgroup of A. This subgroup contains necessarily d, b, c so that the sum $c+d$ exists in A. Thus we have verified the validity of XI, (c); and the validity of XI, (b) is shown likewise. Hence XI is true; and it follows from Theorem 1 that (T) is satisfied by A. We have shown therefore that (i) is a consequence of (iii); and this completes the proof.

Remark. We note that adds A which satisfy the conditions of Theorem 2 satisfy the strong associative law and are amalgams in the strict sense, as was shown during the proof. Thus they belong to a class of adds, investigated in I, §6.

Example. Form the direct sum $H \oplus J \oplus K \oplus L$ of groups H, J, K, L all different from 0. Form the subadd A of $H \oplus J \oplus K \oplus L$ which consists of the elements in $H \oplus J$, $J \oplus K$, $K \oplus L$. It is a consequence of II, §4, Corollary 2 that A is an amalgam in the strict sense which satisfies the strong associative law. But it follows from Theorem 2 that the associative law (T) is not satisfied, since Condition (D) does not hold in A.

2. The similarity problem. A solution of the similarity problem consists in devising a definite method for deciding in a finite number of steps whether two given vectors are similar or not. This problem is often referred to as the identity problem, since similar vectors represent the same element in the derived manifold [Baer (1), §2], and also as the word problem, since the term "word" is often used instead of our "vector."

Every vector possesses irreducible reductions. If the vector v has length n, then one may obtain each of its irreducible reductions in less than n steps. The procedure may be normalized, if we care to do so, by requiring to make always that contraction which is "farthest left." Likewise it would be possible to obtain all irreducible reductions of v in less than $\frac{1}{2}n(n-1)$ steps. Since a vector and its reductions are similar vectors, it follows that it suffices to solve the similarity problem for irreducible vectors. We are going to offer here a solution of this problem for adds which satisfy Postulates I, II and the associative law (T). This solution is essentially contained in the following proposition.

THEOREM. *If Postulates* I, II *and the associative law* (T) *are satisfied by the add A, then the following conditions are necessary and sufficient for similarity of the irreducible vectors $(a_1, \cdots, a_m)$ and $(b_1, \cdots, b_n)$.*

(a) $m = n$.

(b′) *If* $1 = m = n$, *then* $a_1 = b_1$.

(b″) *If* $1 < m = n$, *then there exist elements $c_2, \cdots, c_n$ in A such that*

$$c_2 = -b_1 + a_1,$$
$$c_i = -b_{i-1} + (c_{i-1} + a_{i-1}) = (-b_{i-1} + c_{i-1}) + a_{i-1} \text{ for } 2 < i \leq n,$$
$$c_n = b_n - a_n.$$

Proof. It is a consequence of §**1**, Theorem 1 that Postulates VI, VII are satisfied by A. Consequently Postulates I to VII are satisfied by A; and we may use the operation $-x$ [II, §1].

Assume now the validity of our conditions. If $m = n = 1$, then our vectors are equal [and consequently similar]. If $1 < m = n$, then it follows from our conditions [and $y = -x + (x + y)$] that

$$a_1 = b_1 + c_2,\ c_{i-1} + a_{i-1} = b_{i-1} + c_i \text{ for } 2 < i \leq n, b_n = c_n + a_n.$$

This implies

$$\begin{aligned}(a_1, \cdots, a_m) &\sim (a_1, -c_2, c_2, a_2, \cdots, a_{i-1}, -c_i, c_i, a_i, \cdots, -c_n, c_n, a_n)\\ &\sim (a_1, -c_2, c_2 + a_2, \cdots, c_{i-1} + a_{i-1}, -c_i, c_i + a_i, \cdots, -c_n, c_n + a_n)\\ &= (b_1 + c_2, -c_2, b_2 + c_3, \cdots, b_{i-1} + c_i, -c_i, b_i + c_{i+1}, \cdots, -c_n, b_n)\\ &\sim (b_1, c_2, -c_2, b_2, c_3, \cdots, b_{i-1}, c_i, -c_i, b_i, c_{i+1}, \cdots, -c_n, b_n)\\ &\sim (b_1, \cdots, b_n);\end{aligned}$$

and thus we have shown the sufficiency of our conditions.

Suppose conversely the similarity of the irreducible vectors $(a_1, \cdots, a_m)$

and $(b_1, \cdots, b_n)$. Then $m = n$ is a consequence of the invariance property (L). If $1 = m = n$, then it follows from (S. 1) that $a_1 = b_1$. Thus we have shown the necessity of our condition (b) in case the similar irreducible vectors are of length 1. Hence we asume now that the irreducible similar vectors are of equal length $n > 1$ and that the validity of (b) has been verified for similar irreducible vectors of length less than n. From the similarity of $(a_1, \cdots, a_n)$ and $(b_1, \cdots, b_n)$ we deduce now that

$$\begin{aligned}(a_1, \cdots, a_{n-1}) &\sim (a_1, \cdots, a_{n-1}, a_n, -a_n) = (a_1, \cdots, a_n) + (-a_n)\\ &\sim (b_1, \cdots, b_n) + (-a_n) = (b_1, \cdots, b_n, -a_n).\end{aligned}$$

Since the first of these vectors has length $n - 1$ and the last one has length $n + 1$, it follows from (L) that the last vector is reducible. Since $(b_1, \cdots, b_n)$ is irreducible, this shows the existence of

$$c_n = b_n - a_n.$$

Hence $(a_1, \cdots, a_{n-1}) \sim (b_1, \cdots, b_{n-1}, c_n)$; and it follows as before that the sum $b_{n-1} + c_n$ exists in A. Consequently we have

$$(a_1, \cdots, a_{n-1}) \sim (b_1, \cdots, b_{n-2}, b_{n-1} + c_n).$$

The first of these vectors of length $n - 1$ is irreducible by hypothesis and thus we may deduce the irreducibility of the second vector from the invariance rule (L). Upon these similar irreducible vectors of length $n - 1$ the inductive hypothesis may be applied. Consequently either $n - 1 = 1$ and $a_1 = b_1 + c_2$ or $c_2 = -b_1 + (b_1 + c_2) = -b_1 + a_1$, or else $1 < n - 1$ and there exist elements $c_2, \cdots, c_{n-1}$ in A such that $c_2 = -b_1 + a_1$,

$$c_i = -b_{i-1} + (c_{i-1} + a_{i-1}) = (-b_{i-1} + c_{i-1}) + a_{i-1} \text{ for } 2 < i < n,$$

$c_{n-1} = (b_{n-1} + c_n) - a_{n-1}$. From the last of these equations we deduce first that

$$\begin{aligned}b_{n-1} + c_n &= [(b_{n-1} + c_n) - a_{n-1}] + a_{n-1} = c_{n-1} + a_{n-1},\\ c_n &= -b_{n-1} + (b_{n-1} + c_n) = -b_{n-1} + (c_{n-1} + a_{n-1}).\end{aligned}$$

We have furthermore

$$\begin{aligned}(b_1, \cdots, b_{n-1}) &\sim (b_1, \cdots, b_{n-1}, c_n, -c_n)\\ &\sim (b_1, \cdots, b_{n-2}, b_{n-1} + c_n) + (-c_n)\\ &\sim (a_1, \cdots, a_{n-1}) + (-c_n) = (a_1, \cdots, a_{n-1}, -c_n);\end{aligned}$$

and now we deduce as usual from (L) the existence of $a_{n-1} - c_n$. This implies the existence of $-(a_{n-1}-c_n)=c_n-a_{n-1}$; and consequently it follows from Posulates VI, VII that

$$c_{n-1}=(b_{n-1}+c_n)-a_{n-1}=b_{n-1}+(c_n-a_{n-1}),$$
$$c_n-a_{n-1}=-b_{n-1}+[b_{n-1}+(c_n-a_{n-1})]=-b_{n-1}+c_{n-1},$$
$$c_n=(c_n-a_{n-1})+a_{n-1}=(-b_{n-1}+c_{n-1})+a_{n-1};$$

and this clearly completes the inductive proof of (b).

Remark 1. This theorem contains a solution of the similarity problem, since it gives a definite procedure for finding elements $c_2,\cdots,c_n$—if they cannot be determined in this fashion, then the vectors are not similar—and since it subjects the last element c_n to a further test.

Remark 2. In case A happens to be an amalgam satisfying (T), then one may prove easily the following fact: If C is a class of irreducible similar vectors over A, then all the vectors in C have the same length $n(C)$; and there exists a subgroup $S(i)$ of A which contains all the i-th coordinates of vectors in C.

3. Equality of similar irreducible vectors. If similar irreducible vectors are necessarily equal, then the similarity problem, discussed in §2, has a trivial solution. We are going to characterize these adds.

THEOREM 1. *The following properties of the add A are equivalent.*

(i) *Similar irreducible vectors are equal.*

(ii) *A satisfies Postulates* VI *and*

XII. *If the sums $a+b$, $b+c$ exist in A, though at least one of the sums $a+(b+c)$ and $(a+b)+c$ does not exist, then $a=a+b$ and $c=b+c$.*

(iii) *All contractions of the vector v are connected.*

Proof. Assume first that similar irreducible vectors are equal. If the sums $a+b$ and $b+c$ exist in A, then we have

$$(a+b,c)\sim(a,b,c)\sim(a,b+c).$$

Since the first and last of these vectors are of length 2, it follows that they are either both reducible or both irreducible. If they are both reducible, then we have $((a+b)+c)\sim(a+(b+c))$; and this implies $(a+b)+c=a+(b+c)$. If they are both irreducible, then they are equal, and hence $a+b=a$, $c=b+c$. This shows that (ii) is a consequence of (i).

Assume next the validity of (ii). Then one deduces readily the validity of Postulates VI, VII, XI; and it follows from §1, Theorem 1 that the contractions v' and v'' of the vector v are both irreducible, if they are not connected. It follows then from §1, Lemma that

$$v = u + (a, b, c) + w, \quad v' = u + (a + b, c) + w, \quad v'' = u + (a, b + c) + w,$$

where neither of the sums $(a + b) + c$ and $a + (b + c)$ exists in A. Apply XII to show that $a = a + b$, $c = b + c$ and that therefore $v' = v''$. This is a contradiction, since we assumed that v' and v'' are not connected. Hence we have shown that (iii) is a consequence of (ii).

Assume finally the validity of (iii). Then we are going to show first that

(E′) *every vector possesses one and only one irreducible reduction.*

This is certainly true for vectors v of length 1. Hence we assume that v is a vector of length $n > 1$; and that vectors of length less than n possess one and only one irreducible reduction. There is nothing to prove, if v is irreducible. Hence assume that v be a reducible vector of length n; and that w', w'' are irreducible reductions of v. Then there exist contractions v' and v'' of v such that w' and w'' are reductions of v' and v'' respectively. Then either $v' = v''$, or there exist contractions $v(1), \cdots, v(k)$ of v such that $v' = v(1)$, $v(k) = v''$ and such that $v(i)$ and $v(i + 1)$ possess a common contraction. It follows from the inductive hypothesis that v', v'' and the $v(i)$—as vectors of length $n - 1$—possess one and only one irreducible reduction. This implies $w' = w''$ in case $v' = v''$; and if $v' \neq v''$, then $v(i)$ and $v(i + 1)$ have the same irreducible reduction, since they possess a common contraction and therefore a common irreducible reduction. Hence it follows also in the later case that $v' = v(1)$ and $v'' = v(k)$ possess the same irreducible reduction, namely $w' = w''$. This completes the inductive proof of (E′); and from (E′) we deduce (i) by the customary arguments. This completes the proof.

THEOREM 2. *The following properties of the add A are equivalent.*

(i) *A is an amalgam such that similar irreducible vectors are equal.*

(ii) *A satisfies Postulates* I, II, VI, VII *and*

XII.* *If the sums $a + b$, $b + c$ exist in A, and if $b \neq 0$, then the sum $a + c$ exists in A.*

(iii) *A is an amalgam such that distinct maximal subgroups of A have only* 0 *in common.*

Proof. If (i) is true, then the strong associative law is satisfied by A. Consequently A is an amalgam in the strict sense and satisfies Postulates I to X [II, §3, Theorem & II, §2]. If the sums $a+b$, $b+c$ exist in A, and if $b \neq 0$, then the non-existence of one of the sums $a+(b+c)$ and $(a+b)+c$ would imply [by Theorem 1, XII] that $a=a+b$ and $c=b+c$; and this would imply $b=0$, since a and b as well as b and c belong to some subgroup of A. Consequently the sums $a+(b+c)$, $(a+b)+c$ exist in A; and this implies as usual that a, b, c belong to some subgroup of A. Hence $a+c$ exists in A. Thus (ii) is a consequence of (i).

If (ii) is true, then one verifies readily the validity of Postulates I, II, VI, VII and XI*; and it follows from §1, Theorem 2 that A is an amalgam in the strict sense which satisfies the associative law (T). If X and Y are distinct maximal subgroups of A, then there exists an element x in X and and an element y in Y such that the sum $x+y$ does not exist in A [II, §2, Corollary 1, (c)]. If z belong to the cross cut of X and Y, then the sums $x+z$, $z+y$ exist in A, though $x+y$ does not exist in A. Hence we deduce from XII* that $z=0$ or $X \cap Y=0$. This shows the validity of (iii) so that (iii) is a consequence of (ii).

If finally (iii) is satisfied, and if the sums $a+b$, $b+c$ exist in A, though the sums $a+(b+c)$ and $(a+b)+c$ do not both exist in A, then there does not exist a subgroup of A which contains a, b, c. Consequently there are distinct maximal subgroups X, Y such that a, b are in X and b, c in Y. Then b is in $X \cap Y=0$ so that $b=0$. This shows the validity of Postulate XII; and we deduce from Theorem 1 that similar irreducible vectors are equal. Hence (i) is a consequence of (iii). This completes the proof.

Remark. This theorem shows that the free sums of groups are characterized among the free sums with amalgamations by the property that similar irreducible vectors are equal.

4. The transformation problem. Throughout this section we shall be assured of the validity of the Postulates I to V; as a matter of fact we shall mostly, though not always, assume that the adds under consideration are amalgams in the strict sense [so that all the Postulates I to X are valid]. Thus there exists to every element x in the add A a uniquely determined element $-x$ such that $x-x=-x+x=0$; and we shall use the notation: $-(v_1, \cdots, v_n)=(-v_n, \cdots, -v_1)$.

DEFINITION 1. *The vectors v and w over A are equivalent, if there exists a vector t over A such that $t+v \sim w+t$.*

Since the defining similarity may also be written in the form $v \sim -t + w + t$, this amounts to saying that equivalent vectors are "transforms" of each other. Noting this, it follows that similarity implies equivalence, and that equivalence is symmetric, reflexive, transitive and is in agreement with generally accepted terminology.

One proves easily that the vectors v and w are equivalent if, and only if, there exists a vector u such that $v \sim u$ and w arises from u by a cyclic permutation of the coordinates. Thus we are led to the following concepts.

DEFINITION 2. *The vector v is a cyclic reduction of the vector w, if ther exist vectors $t(0), \cdots, t(k)$ such that $w = t(0)$, $t(k) = v$, and such that $t(i+1)$ is either a reduction of $t(i)$ or else is obtained from $t(i)$ by a cyclic permutation of the coordinates of $t(i)$.*

It is clear that every reduction is also a cyclic reduction and that cyclic reductions of v are equivalent to v.

DEFINITION 3. *If all the vectors arising from v by cyclic permutation of coordinates are irreducible, then v is cyclically irreducible.*

Thus $(v_1, \cdots, v_n)$ is cyclically irreducible, if either $n = 1$ or else none of the sums $v_i + v_{i+1}$ for $0 < i < n$ nor the sum $v_n + v_1$ exists in A.

It is easy to see that every vector possesses cyclic reductions which are cyclically irreducible; and that cyclically irreducible vectors do not possess "proper" cyclic reductions.

DEFINITION 4. *The elements a and b are equivalent over A, if there exists a vector v of non-negative length such that v and $-v$ are irreducible vectors and such that (a) is a reduction of $v + (b) - v$.*

If the strong associative law holds in A, then this concept of equivalence is clearly symmetric, reflexive, and transitive. Whether the strong associative law is needed for the validity of this statement, is an open question.

If we substitute in our various "associative laws" equivalence for similarity of vectors and equivalence for equality of elements, cyclic reduction for reduction and cyclical irreducibility for irreducibility, then we obtain a new kind of associative laws. It is our object to prove the rather surprising fact that, at least for amalgams, these new associative laws are equivalent to the original laws they arose from.

If we apply the procedure just outlined to the strong associative law, then we obtain the following property.

THE ASSOCIATIVE LAW (S_t) consists of the following two properties:

(S_t. 1) *If the vectors* (a) *and* (b) *are equivalent, then the elements* a *and* b *are equivalent over* A.

(S_t. 2) *If the cyclically irreducible vectors* v *and* w *are equivalent, and if one of them has length* 1, *then both have length* 1.

The second property will usually be used in the following form: If (a) and v are equivalent, and if the length of v is greater than 1, then v is cyclically reducible.

THEOREM 1. *The strong associative law and the associative law* (S_t) *are equivalent properties of amalgams.*

This theorem will be a consequence of two slightly stronger propositions.

THEOREM 1a. *The associative law* (S_t) *is a consequence of the strong associative law and the Postulates* I, II, IX.

Proof. It is a consequence of the strong associative law and Postulates I, II that Postulates I to V are satisfied by A [see II, §4, Theorem]. If (a) and (b) are equivalent vectors over A, then there exists a vector v of shortest length such that $(a) \sim -v + (b) + v$. It is clear that v and $-v$ are irreducible vectors; and it follows from the strong associative law that (a) is a reduction of $-v + (b) + v$. Hence a and b are equivalent elements over A [Definition 4]; and thus (S_t. 1) is satisfied by A.

Suppose now that the vector $v = (v_1, \cdots, v_m)$ is cyclically irreducible, has length m greater than 1, and is equivalent to a vector of length 1. Then there exists a vector w of minimal length n such that $w + v - w$ is similar to a vector of length 1. One verifies readily that w and $-w$ are irreducible vectors; and one deduces from the strong associative law that $w + v - w$ possesses a reduction of the form (a). It is now easy to show that $w + v - w = p + q$ where the vectors p and q possess reductions (b) and (c) respectively such that $a = b + c$. If the length of p would not exceed n, then we would have $w = p + p'$. Since w is irreducible, so is p. But (b) is a reduction of p so that $p = (b)$. From the existence of the sum $b + c = a$ and Postulate IX we infer the existence of the sum $c + b = d$. Hence (d) is a reduction of $q + p$. Now we see that

$$(d) \sim q + p = p' + v - p' - p + p \sim p' + v - p';$$

and this is impossible, since p' has length at most $n - 1$, and since w was selected as a vector of minimal length transforming v into a vector of length 1. Thus we have shown that the length of p is greater than n; and one verifies likewise that the length of q is greater than n. Consequently there exists an integer i such that

$$0 < i < m, \quad p = w + (v_1, \cdots, v_i), \quad q = (v_{i+1}, \cdots, v_m) - w.$$

From the existence of the sum $a = b + c$ and Postulate IX we infer again the existence of the sum $c + b = d$. Since (b) and (c) are reductions of p and q respectively, we find that (d) is a reduction of $q + p$. Consequently we have

$$\begin{aligned}(d) \sim q + p &= (v_{i+1}, \cdots, v_m) - w + w + (v_1, \cdots, v_i)\\ &\sim (v_{i+1}, \cdots, v_m, v_1, \cdots, v_i) = v'.\end{aligned}$$

It follows from the strong associative law that (d) is a reduction of v'. But v' arises from v by cyclic permutation of coordinates; and this contradicts the assumed cyclical irreducibility of v. Thus we have been led to a contradiction which proves the validity of $(S_t. 2)$. This completes the proof.

Remark. It appears probable that Postulate IX is not necessary for the validity of (S_t), though it may be indispensable.

THEOREM 1b. *The strong associative law is a consequence of the associative law* (S_t) *and the Postulates* I *to* V.

Proof. It is a consequence of II, §4, Theorem that it suffices to prove the validity of the following proposition:

(*) *If* $(0) \sim v$, *then* (0) *is a reduction of* v.

This we prove by complete induction with respect to the length n of the vector v. If firstly $v = (a)$ is of length 1, then we infer from $(0) \sim (a)$ the equivalence of (0) and (a); and we infer from Property $(S_t. 1)$ the equivalence of the elements 0 and a in A. Hence there exists a vector t such that t and $-t$ are irreducible vectors and such that (0) is a reduction of $t + (a) - t$. It follows from II, §4, Lemma 2 that (0) is likewise a reduction of $-t + t + (a)$. We infer from II, 4, Lemma 1, (b) the existence of a reduction (b) of $-t + t$ such that $b + a = 0$. But t and $-t$ are irreducible vectors; and thus (0) is the only reduction of $-t + t$. Hence $b = 0$, and consequently $a = 0$. This proves (*) for vectors v of length 1.

Now we assume that $(0) \sim w$, that w is a vector of length $n + 1$ with $0 < n$, and that we have verified (*) for all vectors v whose length does not exceed n. Similarity implies equivalence; and it follows from Property $(S_t. 2)$ that w is cyclically reducible. We distinguish two cases.

Case 1. w is reducible. Then there exists a contraction w' of w. Since clearly $(0) \sim w \sim w'$, and since w' has length n, we may apply the inductive hypothesis which assures us that (0) is a reduction of w' and hence of w.

Case 2. w is irreducible. Then it follows from the cyclic reducibility of w that $w = (r) + s + (t)$ and that the sum $t + r$ exists in A. We have $(t, r) + s \sim (t) + w - (t) \sim (0)$. Since $(t, r) + s$ is reducible, it follows from Case 1 that (0) is a reduction of $(t, r) + s$. But then it follows from II, §4, Lemma 2 that (0) is likewise a reduction of the vector w which arises from $(t, r) + s$ by a cyclic permutation of coordinates. This completes the induction, the proof of (*) and of Theorem 1b, as has been pointed out before.

That Theorem 1 is an immediate consequence of Theorems 1a and 1b, is fairly obvious, since Postulates I to V and IX are satisfied by all amalgams.

THE ASSOCIATIVE LAW $(\mathrm{T_t})$ consists of the Property $(\mathrm{S_t}.\,1)$ and of the following invariance property:

$(\mathrm{L_t})$ *Equivalent cyclically irreducible vectors have the same length.*

It is clear that $(\mathrm{L_t})$ implies $(\mathrm{S_t}.\,2)$. Hence $(\mathrm{S_t})$ is a consequence of $(\mathrm{T_t})$. If, furthermore, Postulates I to V are satisfied, then it follows from Theorem 1b that the strong associative law is likewise satisfied by A.

The following notation will prove convenient, since we are going to restrict our discussion to amalgams.

DEFINITION 5. *The subset S of the add A is co-group, if there exists a subgroup T of A such that $S \leq T$.*

Co-group subsets are clearly contained in maximal subgroups, since every subgroup is contained in a maximal subgroup.

THEOREM 2. *The associative law* (T) *and the associative law* $(\mathrm{T_t})$ *are equivalent properties of amalgams.*

Proof. We have noted before that the strong associative law is always a consequence of (T); and that in the presence of Postulates I to V [which are satisfied in all amalgams] the strong associative law is likewise a consequence of $(\mathrm{T_t})$. Thus we may assume throughout the proof that the strong associative law is satisfied by the amalgam A; and that therefore A is an amalgam in the strict sense of the word [II, §3, Theorem]. Thus all the postulates I to X are satisfied by A [II, §2] and also the property:

The elements x, y, z are co-group if, and only if, the sums $x + y$, $y + z$, $z + x$ exist in A.

We note furthermore the equivalence of the following three properties of elements a, b in A:

$$a, b \text{ are co-group}; \ a + b \text{ exists in } A; \ a - b \text{ exists in } A.$$

A. Assume the validity of $(\mathrm{T_t})$. Consider elements a, b, c, d in A whose differences $a - b$, $b - c$, $c - d$ exist in A. Then we have the following vector similarities:

$$(a - b, b - c, c - d) \sim (a, -b, b, -c, c, -d) \sim (a, -d).$$

Thus we have shown the similarity, and hence the equivalence, of a vector of length 3 and a vector of length 2. It follows from $(\mathrm{L_t})$ that the vector of length 3 is not cyclically irreducible. Hence at least one of the sums $(a - b) + (b - c)$, $(b - c) + (c - d)$ and $(c - d) + (a - b)$ exists in A.

Case 1. $(a - b) + (b - c)$ exists in A. Then we have the vector similarities:

$$((a - b) + (b - c)) \sim (a, -b, b, -c) \sim (a, -c);$$

and we infer from the strong associative law the existence of the difference $a - c$.

Case 2. $(b - c) + (c - d)$ exists in A. Then one deduces [as in Case 1] the existence of $b - d$.

Case 3. $(c - d) + (a - b)$ exists in A. Then there exists $(a - b) + (c - d)$ and we have

$$((a - b) + (c - d)) \sim (a, -b, c, -d) \sim (a, -b + c, -d).$$

[Note that the existence of $-b + c$ is a consequence of the existence of $c - b$.] It follows from the strong associative law that at least one of the sums $a + (-b + c)$ and $(-b + c) - d$ exists in A. If $a + (-b + c)$ exists in A, then the pairs $a, -b + c$ and $-b + c, -b$ and $a, -b$ are pairs of co-group elements in A. But A is an amalgam in the strict sense so that $a, -b + c, -b$ are co-group. Hence a, b, c are co-group so that $a - c$ exists in A. If $(-b + c) - d$ exists in A, then we show likewise the existence of $b - d$ in A.

Thus we have shown that in every case at least one of the differences $a - c$, $b - d$ exists in A. The validity of (T) is therefore a consequence of §1. Corollary.

B. Assume conversely the validity of (T). Then one infers immediately from §**1**, Theorem 2, (iii) and the fact that A is an amalgam in the strict sense the validity of the following proposition.

(B. 1) *If a, b and b, c and c, d are three pairs of co-group elements in A, then at least one the triplets a, b, c and b, c, d is co-group.*

Next we prove the following fundamental proposition.

(B. 2) *Suppose that v and w are similar irreducible vectors. Then v is cyclically irreducible if, and only if, w is cyclically irreducible.*

It is a consequence of the invariance of length that v and w are vectors of equal length n. Because of the symmetry of our hypotheses it suffices to show that cyclical reducibility of w implies cyclical reducibility of v. The irreducible, cyclically reducible vector w has the form $w = (w_1, \cdots, w_n)$, where none of the sums $w_i + w_{i+1}$ with $0 < i < n$ exists in A, whereas the sum $w_n + w_1$ exists in A, so that in particular $2 < n$. Let $v = (v_1, \cdots, v_n)$. Then we infer from $v \sim w$ and §**2**, Theorem the existence of $c = w_n - v_n$, $d = w_1 - v_1$ and the fact that c, w_{n-1}, v_{n-1} and d, w_2, v_2 are co-group triplets. We have furthermore the following vector similarities:

$$(w_n + w_1, w_2, \cdots, w_{n-2}, w_{n-1} + c) \sim (w_n, w_1, \cdots, w_{n-1}, c)$$
$$\sim (w_n) + w + (-v_n) \sim (w_n) + v + (-v_n) \sim (w_n, v_1, \cdots, v_{n-1}).$$

Since the first of these vectors has length $n - 1$, whereas the length of the last one is n, it follows from the invariance of length that the last of these vectors is reducible; and this implies because of the irreducibility of v the existence of the sum $w_n + v_1$. Thus we have seen that the three pairs w_n, v_1 and v_1, w_1 and w_1, w_n are co-group. But A is an amalgam in the strict sense; and thus it follows that w_n, w_1, v_1 are co-group. Interchanging 1 and n it follows that w_1, w_n, v_n are co-group too. Hence there exist maximal subgroups X, Y, Z of A such that v_n, w_n, w_1 are in X, v_1, w_1, w_n in Y and d, w_2, v_2 in Z. If $X \cap Y \leq Y \cap Z$, then w_1, w_n would be in Z. But w_2 is in Z and w is irreducible. Hence $X \cap Y \not\leq Y \cap Z$; and it follows from §**1**, Theorem 2 that $Z \cap Y \leq Y \cap X$. But $d = w_1 - v_1$ belongs to $Z \cap Y$ and hence to X. Hence $w_1 - v_1, v_n, w_1$ are in X; and this implies the existence of $v_n + v_1$ so that v is cyclically reducible too, as we claimed.

(B. 3) *If w is a cyclically irreducible vector of length $m > 1$, if $z = (z_1, \cdots, z_n)$ is an irreducible vector of positive length, and if $z + w - z$ is similar to a cyclically irreducible vector, then $(z_n) + w - (z_n)$ is similar to a cyclically irreducible vector of length m.*

As a matter of fact we are going to prove slightly more, namely that $(z_n) + w + (-z_n)$ possesses a cyclically irreducible reduction of length m. The proof of this fundamental proposition will be effected in a number of steps.

(a) *If the sums $a + b$, $(a + b) + c$, $[(a + b) + c] + d$ exist in A, then at least one of the triplets a, b, c and $a + b, c, d$ is co-group.*

It follows from our hypotheses that the pairs a, $a + b$ and $a + b$, $(a + b) + c$ and $(a + b) + c, d$ are co-group; and hence it follows from (B. 1) that at least one of the triplets a, $a + b$, $(a + b) + c$ and $a + b$, $(a + b) + c, d$ is co-group. In the first case a, b, c are co-group; and in the second case $a + b, c, d$ are co-group.

(b) *Suppose that $w = (w_1, \cdots, w_n)$ is a cyclically irreducible vector and that $1 < n$.*

(b. 1) *If the sums $z + w_1$, $(z + w_1) + w_2$ exist in A, then $([z + w_1] + w_2, w_3, \cdots, w_n, -z)$ is a cyclically irreducible vector.*

(b. 2) *If the sums $w_n - z$, $w_{n-1} + (w_n - z)$ exist in A, then $(z, w_1, \cdots, w_{n-2}, w_{n-1} + [w_n - z])$ is a cyclically irreducible vector.*

(b. 3) *If the sums $z + w_1, w_n - z$ exist in A, then $(z + w_1, w_2, \cdots, w_{n-1}, w_n - z)$ is a cyclically irreducible vector.*

If the sums $z + w_1$, $(z + w_1) + w_2$ exist in A, then the sum $[(z + w_1) + w_2] + w_3$ cannot exist in A [this is vacuous, if $n = 2$], since it would otherwise follow from (a) that one of the triplets z, w_1, w_2 and $z + w_1, w_2, w_3$ is co-group, which would contradict the irreducibility of w. Furthermore neither $-z + [(z + w_1) + w_2]$ nor $[(z + w_1) + w_2] - z$ can exist in A, since otherwise both these sums would exist in A, and since this would imply by (a) that at least one of the triplets z, w_1, w_2 and $z + w$, $w_2, -z$ is co-group. But in either case $w_1 + w_2$ would exist, which is impossible. This proves already (b. 1) in case $n = 2$. To finish the proof of (b. 1) we have to show that $w_n - z$ cannot exist in case $2 < n$. But if $w_n - z$ existed in A, then the pairs w_n, z and $z, z + w_1$ and $z + w_1, w_2$ would be co-group. This would imply by (B. 1) that at least one of the triplets $w_n, z, z + w_1$ and $z, z + w_1, w_2$ would be co-group. In the first case $w_n + w_1$ would exist in A; and in the second case $w_1 + w_2$. But this contradicts the cyclical irreducibility of w. This completes the proof of (b. 1); and the validity of (b. 2) is shown likewise [or may be inferred by reasons of symmetry].

To prove (b. 3) assume the existence of $z + w_1$ and $w_n - z$. Then it follows from (b. 1) that $(z + w_1) + w_2$ does not exist in A; and it follows from (b. 2) that $w_{n-1} + (w_n - z)$ does not exist in A. The existence of $(w_n - z) + (z + w_1)$ would imply finally that z, w_1, w_n are co-group, which is impossible, since w is cyclically irreducible; and this completes the proof of (b).

Now we are ready to prove (B. 3). Since z is a vector of positive length, $z + w - z$ is not cyclically irreducible. But by hypothesis $z + w - z$ is similar to a cyclically irreducible vector. Hence it follows from (B. 2) that $z + w - z$ is reducible. Since w and z, and consequently also $-z$, are irreducible vectors, it follows that at least one of the sums $z_n + w_1$ and $w_m - z_n$ exists in A. [Here we assume that $w = (w_1, \cdots, w_m)$.] If both $z_n + w_1$ and $w_m - z_n$ exist in A, then it follows from (b. 3) that $(z_n + w_1, w_2, \cdots, w_{m-1}, w_m - z_n)$ is a cyclically irreducible reduction of $(z_n) + w + (-z_n)$. If exactly one of the two sums exists in A, then we may assume wtihout loss in generality that $z_n + w_1$ exists in A, whereas $w_m - z_n$ does not exist in A. If $(z_n + w_1) + w_2$ exists in A, then it follows from (b. 1) that $([z_n + w_1] + w_2, w_3, \cdots, w_m, -z_n)$ is a cyclically irreducible reduction of $(z_n) + w + (-z_n)$. Thus our proof of (B. 3) will be complete, once we have shown that it is impossible to assume that $z_n + w_1$ exists in A, whereas neither $(z_n + w_1) + w_2$ nor $w_m - z_n$ exists in A. If these assumptions were satisfied, then the vector

$$t = (z_n + w_1, w_2, \cdots, w_m, -z_n, -z_{n-1}, \cdots, -z_1)$$

would be irreducible, since w and $-z$ are irreducible vectors. We note the similarity

$$z + w - z \sim (z_1, \cdots, z_{n-1}) + t$$

so that both these vectors are similar to a cyclically irreducible vector.

If $n = 1$, then we infer from (B. 2) that t is cyclically irreducible which contradicts, however, the fact that $-z_1$ and $z_1 + w_1$ are co-group.

If $n = 2$, then it is clear that $(z_1) + t$ is not cyclically irreducible. Since this vector is similar to a cyclically irreducible vector, it follows from

(B. 2) that $(z_1) + t$ is reducible. Hence we may infer from the irreducibility of t the existence of $z_1 + (z_2 + w_1)$. Consequently

$$t^0 = (z_1 + (z_2 + w_1), w_2, \cdots, w_m, -z_2, -z_1) \sim (z_1) + t \sim z + w - z$$

is similar to a cyclically irreducible vector, though it is clear that t^0 itself is not cyclically irreducible, since $-z_1$ and $z_1 + (z_2 + w_1)$ are co-group. It follows from (B. 2) that t^0 is reducible; and this implies that w_2 and $z_1 + (z_2 + w_1)$ are co-group, since t is irreducible. If we interchange right and left in (a)—as we may—then this implies that at least one of the triplets $w_2, z_1, z_2 + w_1$ and z_1, z_2, w_1 is co-group. But this is impossible, since we assumed the irreducibility of z and the non-existence of the sum $(z_2 + w_1) + w_2$.

Thus we may assume finally that $2 < n$. Then we see exactly as in the preceding paragraph of the proof that $z_{n-1} + (z_n + w_1)$ exists in A. Consequently

$$t' = (z_1, \cdots, z_{n-2}, z_{n-1} + (z_n + w_1), w_2, \cdots, w_m) - z \sim z + w - z$$

is similar to a cyclically irreducible vector, though t' is clearly not cyclically irreducible itself. Hence it follows from (B. 2) that t' is reducible; and it follows from the irreducibility of z and w that at least one of the sums

$$z_{n-2} + [z_{n-1} + (z_n + w_1)] \text{ and } [z_{n-1} + (z_n + w_1)] + w_2$$

exists in A. As before, we may now deduce from (a) that at least one of the following three triplets is co-group:

$$z_{n-2}, z_{n-1}, z_n + w_1 \text{ and } z_{n-1}, z_n, w_1 \text{ and } w_2, z_{n-1}, z_n + w_1;$$

and this contradicts the assumed irreducibility of the vectors z and t. Thus we have been led to contradictions in all cases; and this completes the proof of (B. 3).

We are now ready to deduce the associative law (T_t) from (T). We note first that (T) implies the strong associative law; and this implies [by Theorem 1] the associative law (S_t). Thus Property $(S_t. 1)$ is satisfied by A. It follows furthermore from $(S_t. 2)$ that if one of two equivalent, cyclically irreducible vectors has length 1, then both have length 1. Suppose now that v and w are equivalent and cyclically irreducible vectors neither of which has length 1. Then there exists an irreducible vector z of length n such that $z + w - z \sim v$. If z has length 0, then w and v are similar irreducible vectors; and the equality of their length is a consequence of (T). If z has positive length, then $z = z' + (s)$; and it follows from (B. 3) that $(s) + w - (s)$ is similar to a cyclically irreducible vector w' whose length equals the length of w. But then $v \sim z' + w' - z'$ where z' is an irreducible vector of length $n - 1$; and where w' is a cyclically irreducible vector of the same length m as w. By an obvious induction one proves now that v, w' and w have the same length m. Thus (L_t) is true; and this completes the proof of Theorem 2.

Remark. The lemma (b) which we deduced during the proof of the preceding theorem contains a sort of procedure for deriving equivalent cyclically irreducible vectors from a given cyclically irreducible vector; but it does not seem possible to derive from it a straightforward method which could be used to determine whether or not two given cyclically irreducible vectors are equivalent.

THEOREM 3. *Similarity of irreducible vectors over the amalgam A implies their equality if, and only if, A has the following properties:*

(1) *Equivalence of vectors (a) and (b) implies that a and b are conjugate elements in some subgroup of A.*

(2) *Equivalence of the cyclically irreducible vectors v and w which are not both of length* 1 *implies that v arises from w by a cyclic permutation of the coordinates.*

The proof of this theorem will be omitted, since it is not very difficult. It is best based on §3, Theorem 2, (iii).

UNIVERSITY OF ILLINOIS,
URBANA, ILLINOIS.

BIBLIOGRAPHY.

Reinhold Baer.

(1) "Free sums of groups and their generalizations. An analysis of the associative law," *American Journal of Mathematics*, vol. 71 (1949), pp. 706-742.

(2) "Free sums of groups and their generalizations II," *ibid.*, vol. 72 (1950), pp. 625-646.

GENERALIZED FREE SUMS OF CYCLICAL GROUPS.*

By HANNA NEUMANN.

1. Introduction. The well-known clover knot group (i. e. the fundamental group of the residual space in S_3 of the clover knot) is generated by two elements a and b satisfying the relation $a^2 = b^3$. More generally, any torus knot group is generated by two elements a and b with the defining relation $a^m = b^n$. In such a group both a and b are of infinite order, and the intersection of the two cycles generated by a and b respectively is exactly the cycle generated by the common power $a^m = b^n$, no more. Thus, in Schreier's terminology, the torus knot group is the free product of two infinite cycles with one amalgamated subgroup.

We consider the following generalization: A group is generated by elements $a_1, a_2, \cdots$, finite or infinite in number, with the defining relations

$$1.1. \qquad a_i^{m_i} = 1 \qquad (i = 1, 2, \cdots),^1$$

$$1.2. \qquad a_i^{e_{ij}} = a_j^{e_{ji}} \qquad (i \neq j; i, j = 1, 2, \cdots).$$

This group may, or may not, have the following property:

1.3. *All generators a_i are of exact order m_i, and any two generators a_i and a_j have exactly the power $a_i^{e_{ij}} = a_j^{e_{ji}}$ in common, no more.*

Examples of such groups not having this property are easily made. In fact there are some fairly obvious conditions for the numbers e_{ij} which are necessary for the group to satisfy 1.3. They merely express the fact that the relations 1.2 are 'compatible' (in a sense to be precisely defined in **2**) with the relations 1.1 and with each other.

It is the main result of this note that these conditions are also sufficient.

In fact, we shall show more: The compatibility conditions ensure that the group has property 1.3 even if the further relations $a_i a_j = a_j a_i$ are added to 1.1 and 1.2, i. e. if the group is made abelian.

Although the theorem is stronger, its proof is easier: it consists in the explicit construction of a group with the required properties.

In **2** we discuss the problem in a more general context. It arises out of questions on the existence of the 'generalized free product with amalgamated

---

* Received April 1, 1949.

[1] An element of order zero is understood to generate an infinite cycle.

subgroups dealt with in an earlier paper (cf. [4][2]), and closely related to this, R. Baer's investigations on the imbeddability of a certain type of incomplete group called 'amalgam' in a group (cf. [1]). In the terminology of [4] our result proves the existence of the generalized free product of cyclical groups, thus filling a gap left at that time (cf. [4], **10**, p. 624).

In **3** the general question is reduced to the simpler case of a finite number of generators a_i. These may then be assumed to be all of finite, or all of infinite order. The next two paragraphs deal separately with these cases: The compatibility conditions are given in arithmetical form: this allows one to deal with the case of cycles of finite order; the case of infinite cycles is then reduced to the former.

Finally, in **6**, the result is extended to prove the existence of the generalized free sum of *locally* cyclic groups.

I wish to thank Professor Baer and my husband for many helpful remarks, and in particular the referee for a suggestion which led to a substantial simplification of the proof.

2. The context. We recall briefly the definition of a generalized free product with amalgamated subgroups.

Groups $\mathfrak{G}_\alpha$ are given (finite or infinite in number), with subgroups $\mathfrak{U}_{\alpha\beta} \subseteq \mathfrak{G}_\alpha$ such that $\mathfrak{U}_{\alpha\beta}$ and $\mathfrak{U}_{\beta\alpha}$ are isomorphic. Denote by $I_{\alpha\beta}$ a fixed isomorphism mapping $\mathfrak{U}_{\alpha\beta}$ on $\mathfrak{U}_{\beta\alpha}$, and by $I_{\beta\alpha}$ its inverse. If $u_{\alpha\beta}$ is any element of $\mathfrak{U}_{\alpha\beta}$, denote the corresponding element of $\mathfrak{U}_{\beta\alpha}$ by $u_{\beta\alpha}$,

$$u_{\beta\alpha} = I_{\alpha\beta} u_{\alpha\beta} \quad \text{and} \quad u_{\alpha\beta} = I_{\beta\alpha} u_{\beta\alpha}.$$

Introduce in the free product of the groups $\mathfrak{G}_\alpha$ all relations $u_{\alpha\beta} = u_{\beta\alpha}$. If in the resulting group the groups $\mathfrak{G}_\alpha$ remain 'intact,' i. e. isomorphically represented, such that moreover the intersection of $\mathfrak{G}_\alpha$ and $\mathfrak{G}_\beta$ is exactly $\mathfrak{U}_{\alpha\beta} = \mathfrak{U}_{\beta\alpha}$, then this group is called the generalized free product $\mathfrak{G}$ of the groups $\mathfrak{G}_\alpha$ with amalgamated subgroups $\mathfrak{U}_{\alpha\beta}$.

Clearly, for this construction to be successful, the subgroups $\mathfrak{U}_{\alpha\beta}$ and the isomorphisms $I_{\alpha\beta}$ must fit together; more precisely:

2.1. *For the existence of the generalized free product it is necessary that the three groups*

$$\mathfrak{U}_{\alpha\beta} \cap \mathfrak{U}_{\alpha\gamma} = \mathfrak{U}_{\alpha\beta\gamma} \text{ in } \mathfrak{G}_\alpha,$$
$$\mathfrak{U}_{\beta\gamma} \cap \mathfrak{U}_{\beta\alpha} = \mathfrak{U}_{\beta\gamma\alpha} \text{ in } \mathfrak{G}_\beta,$$
$$\mathfrak{U}_{\gamma\alpha} \cap \mathfrak{U}_{\gamma\beta} = \mathfrak{U}_{\gamma\alpha\beta} \text{ in } \mathfrak{G}_\gamma$$

are isomorphic, such that, moreover,

---

[2] Numbers in square brackets refer to the list of references at the end of the paper.

2.2. $I_{\beta\gamma}(I_{\alpha\beta}u_{\alpha\beta\gamma}) = I_{\alpha\gamma}u_{\alpha\beta\gamma}$ *for every element* $u_{\alpha\beta\gamma}$ *of* $\mathfrak{U}_{\alpha\beta\gamma}$.

These conditions are in general not sufficient for the existence of $\mathfrak{G}$ (cf. [4], 3.0 and 3.2, pp. 595, 596).

We return to the group defined in **1.** Property 1.3 is equivalent with the existence of the generalized free product of the cyclical groups (a_i) of order m_i with amalgamated subgroups $(a_i^{e_{ij}})$. The necessary conditions 2.1 and 2.2 obviously become arithmetical conditions for the numbers m_i and e_{ij}—the 'compatibility conditions' mentioned in the introduction—which in this case will be shown to be also sufficient for the existence of the generalized free product. The explicit arithmetical form of these conditions does not concern us until later (**4**, **5**). Instead we consider the conditions 2.1 and 2.2 from a different angle.

A system S of elements is called an incomplete group if a multiplication is defined such that any ordered pair of elements has at most one product, right-division and left-division each have at most one solution, and as far as multiplication is defined it is associative.

Let S consist in particular of all the elements of all the groups $\mathfrak{G}_\alpha$, with amalgamations $u_{\alpha\beta} = u_{\beta\alpha}$ carried out, and with multiplication defined as in the individual groups for any two elements which belong to one and the same group $\mathfrak{G}_\alpha$, and undefined otherwise. This system S is an incomplete group if, and only if, the conditions 2.1 and 2.2 are satisfied (cf. [4], **4**, p. 597).

This type of incomplete group we call, with R. Baer, an 'amalgam' (cf. [1], **6**, p. 728). The groups $\mathfrak{G}_\alpha$ we call its 'constituents.'

Remark. One and the same amalgam may be representable as an amalgam of groups in different ways, even involving a finite number of groups one way, but infinitely many the other way. We are, however, interested not in the amalgam as abstract algebraic system, but in given groups formed into an amalgam by virtue of given amalgamations.

The generalized free product of the groups $\mathfrak{G}_\alpha$ may now be described as the group 'freely generated' by the amalgam of these groups. It exists if, and only if, the amalgam is imbeddable in a group (cf. [4], 4.0, p. 598).

As mentioned already, not every amalgam is imbeddable. R. Baer even gives an example of an amalgam of four abelian groups which is not imbeddable (cf. [1], p. 728). Our result will show that every amalgam of cyclical groups is imbeddable.

Grace E. Bates has generalized the above concepts as follows (cf. [2], Appendix, for more precise definitions): Let all the groups in the amalgam have property P. The generalized P-free product—if it exists—is the group with property P which contains and is generated by the amalgam, and is such that every other group with the same properties is a homomorphic image of it. If, in particular, property P is taken as 'being abelian,' we are led to the generalized free sum $\mathfrak{A}$ of the abelian groups $\mathfrak{A}_a$. Again, $\mathfrak{A}$ exists if, and only if, the amalgam A of the groups $\mathfrak{A}_a$ is imbeddable in an abelian group.

Clearly, the existence of the generalized free sum implies the existence of the generalized free product. Whether the converse is true in general, is an unsolved problem.[2a] In our special case however it is true: As we are going to imbed any given amalgam of cyclical groups in an *abelian* group, we prove the existence of the generalized free sum, as well as that of the generalized free product, of cyclical groups.

3. Preliminaries. In an amalgam of given groups any subsystem of these groups forms a subamalgam. R. Baer shows that:

An amalgam of an infinite number of groups is imbeddable if every subamalgam formed by a finite number of its constituents is imbeddable. In other words:

Imbeddability of an amalgam is a 'property of finite character.'

It follows similarly:

3. 0. *Imbeddability of an amalgam of abelian groups in an abelian group is a property of finite character.*

The proof is immediate and is omitted.[3]

DEFINITION. *The amalgam A with constituent groups $\mathfrak{G}_a$ is called reducible if there exist in A two proper subamalgams A_1 and A_2 with the following properties:*

(i) *neither A_1 nor A_2 consists of the unit element only;*

(ii) *every constituent $\mathfrak{G}_a$ of A belongs to either A_1 or A_2, but not to both;*

(iii) *every $\mathfrak{G}_a$ in A_1 has unit meet with every $\mathfrak{G}_\beta$ in A_2.*

A is called irreducible if no such subamalgams exist in A. An amalgam consisting of one group only is trivially irreducible.

[2a] [Added in proof, September 1950.] This question has now been settled: I have constructed an amalgam of five abelian groups, imbeddable in a group, but not imbeddable in an *abelian* group. (To appear elsewhere.)

[3] Let Σ be a finite set of identical relations. Then one proves equally easily: Imbeddability of an amalgam of groups whose elements satisfy the relations of Σ in another such group is a property of finite character.

Remark. Reducibility of an amalgam A is actually seen to be independent of the particular way in which A is represented as amalgam of constituent groups. We shall, however, not make use of this fact.

3. 11. *Let A be a reducible amalgam of a finite number of abelian constituents, and let A_1 and A_2 be subamalgams satisfying the conditions* (i)-(iii) *above. If A_1 and A_2 are imbeddable in abelian groups $\mathfrak{A}_1$ and $\mathfrak{A}_2$ respectively, then A is imbeddable in an abelian group, viz. the direct sum $\mathfrak{A}$ of $\mathfrak{A}_1$ and $\mathfrak{A}_2$.*

The proof is obvious. It now follows immediately:

3. 12. *A reducible amalgam of a finite number of abelian constituents is imbeddable in an abelian group if this is true for every irreducible subamalgam.*

Applied to our original problem, 3. 0 and 3. 12 show that we may restrict ourselves to irreducible amalgams formed by a finite number of cycles. But then clearly, either all cycles in the amalgam are of finite order, or they are all of infinite order.

3. 2. *Let A be an amalgam of a finite number of cycles which are either all infinite, or all finite and their orders powers of one and the same prime. If A is irreducible, any two cycles in A have non-trivial meet.*

Proof. Under the assumptions of 3. 2 every constituent group of A has the property that any two non-trivial subgroups have non-trivial meet. Hence 3. 2 follows from the definition of irreducibility in conjunction with 2. 1.

Remark. It follows from a result by R. Baer (cf. [1], **9**, p. 742) that an amalgam of cycles whose orders are all powers of one and the same prime is imbeddable. The case of an amalgam of cycles of arbitrary finite orders can be reduced to this case by splitting up each cycle into a direct sum of cycles of prime power orders. As we wish to prove imbeddability in an abelian group, and besides are interested also in the conditions for given amalgamations 1. 2 to define an amalgam, we shall not rely on Baer's result.

To provide a basis for induction, we note:

3. 31. *Every amalgam of two abelian constituents is imbeddable in an abelian group.*

In fact, more can easily be shown; in obvious terminology (cf. [3], **2**):

3. 32. *The direct product of two groups with one amalgamated subgroup exists provided only the amalgamated subgroup belongs to the centre of each group.*

We are now ready to prove the existence of an abelian group containing a given irreducible amalgam of a finite number of cycles. As we shall now be dealing with abelian groups only, we shall write the groups additively.

DEFINITION. *The system of numbers* e_{ij} $(i, j = 1, \cdots, n)$ *is called compatible with elements of order* m_i $(i = 1, \cdots, n)$, *if the relations* $e_{ij}a_i = e_{ji}a_j$ *and* $m_ia_i = 0$ *for all* i *and* $j \neq i$ *define an amalgam of the* n *cycles* (a_i) *of orders* m_i.

3.4. *The system of numbers* e_{ij} *is compatible with elements of order* m_i *if, and only if, the subgroups and amalgamations defined by it satisfy* 2.1 *and* 2.2.

The general pattern of the subsequent proofs is as follows: From the compatibility of the system e_{ij} with elements of order m_i we derive arithmetical conditions satisfied by the numbers e_{ij} and m_i. These *necessary* conditions for compatibility are then shown to imply the existence of an abelian group containing the given amalgam. In this way the arithmetical conditions are shown to be also sufficient, and simultaneously the imbeddability of the amalgam is established.

4. Amalgams of finite cycles. We deal with this case in three steps: first we assume the cycles to be of equal prime power order p^{μ}. Then the case of n cycles of arbitrary prime power orders p^{μ_i} $(i = 1, \cdots, n)$ is reduced to the previous case; finally the case of n cycles of arbitrary finite orders m_i is reduced to the case of cycles of prime power orders.

Throughout the rest of the paper we write $r \vee s$ for the positive greatest common divisor of two integers r and s. Also, if p is any prime, $p^{\lambda(r)}$ the highest power of p dividing the integer r, we call $\lambda(r)$ the p-contents of r, i. e. $r = p^{\lambda(r)}r_1$, where $r_1 \vee p = 1$. (No ambiguity will arise from the fact that this notation $\lambda(r)$ does not indicate the particular prime p under consideration).

Given are n elements a_i with $p^{\mu}a_i = 0$ and $e_{ij}a_i = e_{ji}a_j$. We put

4.01. $\lambda(e_{ij}) = \epsilon_{ij}$, i. e. $e_{ij} = d_{ij}p^{\epsilon_{ij}}$ with $d_{ij} \vee p = 1$.

Since only irreducible amalgams need be considered, we shall assume throughout (because of 3.2):

4.02. $\epsilon_{ij} < \mu$ for all pairs i, j.

4.1. LEMMA. *If the system of numbers* $e_{ij} = d_{ij}p^{\epsilon_{ij}}$ *is compatible with* n *elements of order* p^{μ}, *then*

4.10. $\epsilon_{ij} = \epsilon_{ji}$ for all pairs i, j;

4.11. *of any three exponents* ϵ_{ij}, ϵ_{jk}, ϵ_{ki} *two are equal, the third not greater than these two*;

4.12. *for any three suffixes* i, j, k

$$d_{ij}d_{jk}d_{ki} \equiv d_{ji}d_{kj}d_{ik} \bmod p^{\mu-\epsilon} \textit{ where } \epsilon = \mathrm{Max}(\epsilon_{ij}, \epsilon_{jk}, \epsilon_{ki}).$$

Proof. Since the relations $e_{ij}a_i = e_{ji}a_j$ define an amalgam of the cycles (a_i) of order p^μ, the amalgamated subgroups are isomorphic, which implies 4.10.

To prove 4.11, we use the fact that the amalgamations satisfy 2.1. If for a fixed triplet i, j, k the three exponents ϵ_{ij}, ϵ_{jk}, and ϵ_{ki} are not equal, assume without loss of generality that $\epsilon_{ij} > \epsilon_{ik}$. Thus the amalgamated subgroup $(e_{ij}a_i)$ of (a_i) is contained in $(e_{ik}a_i)$, and their intersection, i. e. the intersection of (a_i), (a_j), and (a_k), is $(e_{ij}a_i)$ of order $p^{\mu-\epsilon_{ij}}$. But this same intersection, formed in (a_k), is the intersection of $(e_{ki}a_k)$ and $(e_{kj}a_k)$. For this to have the same order $p^{\mu-\epsilon_{ij}}$, necessarily $\epsilon_{kj} = \epsilon_{ij}$, because of 4.10. This proves 4.11.

To prove 4.12, we note that the amalgamations satisfy 2.2, i. e. are consistent with each other. Assume $\epsilon_{ij} \geqq \epsilon_{ik}$, and consider the element $p^{\epsilon_{ij}}d_{ij}d_{jk}d_{ki}a_i$ of (a_i). Repeated application of the given relations (noting $p^{\epsilon_{ij}}d_{ij} = e_{ij}$, etc.) gives: $p^{\epsilon_{ij}}d_{ij}d_{jk}d_{ki}a_i = p^{\epsilon_{ij}}d_{ji}d_{jk}d_{ki}a_j$, which is $p^{\epsilon_{ij}}d_{ji}d_{kj}d_{ki}a_k$ (as $\epsilon_{ij} = \epsilon_{jk}$), or $p^{\epsilon_{ij}}d_{ji}d_{kj}d_{ik}a_i$ (as $\epsilon_{ij} \geqq \epsilon_{ik}$); hence these two multiples of a_i represent the same element of (a_i), and 4.12 follows.

4.2. Lemma. *If the system of numbers* e_{ij} *and* p^μ *satisfies the conditions* 4.10-4.12, *then there exists an abelian group* $\mathfrak{A}$ *generated by elements* $a_1, \cdots, a_n$ *of orders* p^μ *such that in* $\mathfrak{A}$ *the intersection of* (a_i) *and* (a_j) *is generated by the element* $e_{ij}a_i = e_{ji}a_j$ *for all* i *and* j.

Proof. If $n = 2$, the lemma is true by 3.3. We shall, therefore, prove it by induction with respect to n.

Assume that $\mathfrak{A}_1$ is an abelian group generated by elements $a_1, \cdots, a_{n-1}$ of orders p^μ, such that in $\mathfrak{A}_1$

4.20. $(a_i) \cap (a_j) = (e_{ij}a_i = e_{ji}a_j)$ for $i, j = 1, \cdots, n-1$.

We construct an abelian group $\mathfrak{A}$ containing the given amalgam of n cycles as free sum of $\mathfrak{A}_1$ and a cycle (a_n) of order p^μ with a suitable amalgamation.

Amongst the intersections (a_n) is to have with the cycles (a_i) $(i \leq n-1)$, there is a maximal one which may be assumed to be its intersection with (a_1), i. e. the subgroup of order $p^{\mu-\epsilon_{n1}}$ generated by $e_{n1}a_n$. The numbers e_{ij} satisfy the conditions 4. 10-4. 12; because of the first of these, the subgroup $(e_{1n}a_1)$ of (a_1), i. e. of $\mathfrak{A}_1$, is isomorphic with $(e_{n1}a_n)$. Therefore the free sum $\mathfrak{A}$ of $\mathfrak{A}_1$ and (a_n) with the amalgamation $e_{1n}a_1 = e_{n1}a_n$ exists. We prove that $\mathfrak{A}$ has the desired properties.

Since $\mathfrak{A}_1$ is isomorphically contained in $\mathfrak{A}$, the intersection of any two of the first $n-1$ cycles (a_i) is given by 4. 20 also in $\mathfrak{A}$. All we have to prove therefore is that also the intersection of (a_i) and (a_n) is the group generated by $e_{in}a_i = e_{ni}a_n$, for $i = 1, \cdots, n-1$. For $i = 1$ this follows from the construction of $\mathfrak{A}$, for $i > 1$ we deduce it from 4. 1.

First we show that $(a_i) \cap (a_n)$ is the subgroup of order $p^{\mu-\epsilon_{in}}$ of (a_n), $(2 \leq i \leq n-1)$.

Now $\mathfrak{A}_1 \cap (a_n) = (e_{1n}a_1 = e_{n1}a_n)$, and $(a_i) \subseteq \mathfrak{A}_1$. Hence

$$(a_i) \cap (a_n) = (a_i) \cap (e_{1n}a_1) \cap (a_n) = (a_i) \cap (e_{1n}a_1),$$

so that

4. 21. $\quad (a_i) \cap (a_n) = (a_i) \cap (a_1)$, if $(a_i) \cap (a_1) \subset (e_{1n}a_1)$,

4. 22. $\quad (a_i) \cap (a_n) = (e_{1n}a_1)$, if $(a_i) \cap (a_1) \supseteq (e_{1n}a_1)$.

But $(a_i) \cap (a_1)$ is of order $p^{\mu-\epsilon_{i1}}$, $(e_{1n}a_1)$ is of order $p^{\mu-\epsilon_{1n}}$. Hence, in the first case (4. 21), $\epsilon_{i1} > \epsilon_{1n}$, so that by 4. 11 $\epsilon_{in} = \epsilon_{i1}$; in the second case (4. 22), $\epsilon_{i1} \leq \epsilon_{1n}$, hence, by 4. 11, $\epsilon_{in} = \epsilon_{1n}$. Thus in both cases $(a_i) \cap (a_n)$ has the desired order.

It remains to prove that the correct multiples $e_{in}a_i = d_{in}p^{\epsilon_{in}}a_i$ and $e_{ni}a_n = d_{ni}p^{\epsilon_{in}}a_n$ of $p^{\epsilon_{in}}a_i$ and $p^{\epsilon_{in}}a_n$ respectively are equal in $\mathfrak{A}$.

In the following lines, d_{ij}^{-1} denotes the unique inverse mod p^{μ} of d_{ij}.

In the first case (4. 21), $(a_i) \cap (a_n)$ is the subgroup $(p^{\epsilon_{in}}a_i)$ of (a_i), which, because of $\epsilon_{in} = \epsilon_{i1}$, is also generated by

$$e_{in}a_i = d_{in}p^{\epsilon_{in}}a_i = d_{in}d_{i1}^{-1}d_{i1}p^{\epsilon_{in}}a_i = d_{in}d_{i1}^{-1}d_{1i}p^{\epsilon_{in}}a_1.$$

Also, as $\epsilon_{in} > \epsilon_{1n}$,

$$e_{ni}a_n = d_{ni}p^{\epsilon_{in}}a_n = d_{ni}d_{n1}^{-1}d_{n1}p^{\epsilon_{in}}a_n = d_{ni}d_{n1}^{-1}d_{1n}p^{\epsilon_{in}}a_1;$$

therefore by 4. 12: $e_{in}a_i = e_{ni}a_n$.

In the second case (4. 22) the proof is similar and will be omitted.

Next we turn to the case of n elements a_i of orders p^{μ_i}, and $e_{ij}a_i = e_{ji}a_j$. The p-contents of e_{ij} is denoted by ϵ_{ij} as before.

4. 3. LEMMA. *If the system of numbers $e_{ij} = d_{ij}p^{\epsilon_{ij}}$ is compatible with n elements of orders p^{μ_i}, and if*

4. 30. $\mu = \operatorname{Max} \mu_i$ *and* $\delta_{ij} = \epsilon_{ij} + \mu - \mu_i,$

then

4. 31. $\delta_{ij} = \delta_{ji}$ *for all pairs* i, j;

4. 32. *of any three numbers δ_{ij}, δ_{jk}, δ_{ki} two are equal, the third not greater than these two;*

4. 33. *for any three suffixes i, j, k:*

$$d_{ij}d_{jk}d_{ki} \equiv d_{ji}d_{kj}d_{ik} \bmod p^{\mu-\delta} \text{ where } \delta = \operatorname{Max}(\delta_{ij}, \delta_{jk}, \delta_{ki}).$$

4. 4. LEMMA. *If the system of numbers e_{ij} and p^{μ_i} satisfies the conditions* 4. 30-4. 33, *then there exists an abelian group $\mathfrak{A}$ generated by elements a_i of orders p^{μ_i} $(i = 1, \cdots, n)$, such that in $\mathfrak{A}$ the intersection $(a_i) \cap (a_j)$ is generated by the element $e_{ij}a_i = e_{ji}a_j$, for all i and j.*

Proof of 4. 3. Consider n elements $b_1, \cdots, b_n$ of equal order p^μ, and in each cycle (b_i) the element $a_i = p^{\mu-\mu_i}b_i$ $(i = 1, \cdots, n)$. Since a_i is of order p^{μ_i}, the relations $e_{ij}a_i = e_{ji}a_j$ define an amalgam of the $a_i = p^{\mu-\mu_i}b_i$, i. e. the relations $e_{ij}p^{\mu-\mu_i} \cdot b_i = e_{ji}p^{\mu-\mu_j} \cdot b_j$ define an amalgam of the cycles b_i. Thus the system of numbers $e_{ij}p^{\mu-\mu_i}$ is compatible with n elements of equal order p^μ, and 4. 3 follows from 4. 1.

Proof of 4. 4. With the same notation as in the proof of 4. 3, we have: under the assumption of 4. 4, the system of numbers $p^{\mu-\mu_i}e_{ij}$ and p^μ satisfies the conditions which, by 4. 2, are sufficient for the existence of an abelian group $\mathfrak{B}$ generated by n elements b_i of equal order p^μ such that $(b_i) \cap (b_j)$ is generated by $e_{ij}p^{\mu-\mu_i}b_i = e_{ji}p^{\mu-\mu_j}b_j$. But then clearly the elements $a_i = p^{\mu-\mu_i}b_i$ of $\mathfrak{B}$ and the subgroup $\mathfrak{A}$ of $\mathfrak{B}$ generated by them satisfy the requirements of the lemma.

Finally, let $a_1, \cdots, a_n$ be elements of arbitrary orders $m_1, \cdots, m_n$, and let again $e_{ij}a_i = e_{ji}a_j$. For any prime p let $\lambda_p(m_i)$ denote the p-contents of m_i.

4. 5. LEMMA. *The system of numbers e_{ij} is compatible with n elements of orders m_i if, and only if, it is compatible with n elements of orders $p^{\lambda_p(m_i)}$ for every prime p.*

4. 6. THEOREM. *Every irreducible amalgam of a finite number of cycles of finite order is imbeddable in an abelian group.*

Proof of 4. 5 *and* 4. 6. Clearly 4. 5 is void for any p which does not divide any of the numbers m_i. Let $p_1, \cdots, p_r$ be all the primes dividing at least one of the m_i. The cycle (a_i) of order m_i is the direct sum of r cycles $(b_{i\rho})$ of order $m_{i\rho}$, where $m_{i\rho} = p_\rho^{\lambda_\rho(m_i)}$, and $\rho = 1, \cdots, r$:

$$a_i = b_{i1} + b_{i2} + \cdots + b_{ir} \text{ with } m_{i\rho} b_{i\rho} = 0,$$

for $i = 1, \cdots, n$. The amalgamations $e_{ij}a_i = e_{ji}a_j$ induce the same amalgamations $e_{ij}b_{i\rho} = e_{ji}b_{j\rho}$ for each fixed value of ρ. If the cycles (a_i) form an amalgam A by means of the given amalgamations, then clearly the cycles $(b_{i\rho})$ (for $i = 1, \cdots, n$) form an amalgam B_ρ by means of the same amalgamations, and conversely. Also, the amalgams B_ρ may be imbedded in abelian groups $\mathfrak{B}_\rho$, by 4. 3 and 4. 4. But then A is imbeddable in the direct sum $\mathfrak{A}$ of the abelian groups $\mathfrak{B}_\rho$.

5. Amalgams of infinite cycles. Again we begin by deriving necessary conditions for the numbers e_{ij} to be compatible with n elements of infinite order.

Given are n elements of infinite order, $a_1, \cdots, a_n$ with $e_{ij}a_i = e_{ji}a_j$. Only irreducible amalgams need be considered, hence we shall assume throughout this paragraph (because of 3. 2).

5. 0. $$e_{ij} \neq 0 \text{ for all } i, j.$$

If the given relations define an amalgam of the n cycles, the subgroups and amalgamations defined by them satisfy 2. 1 and 2. 2. The former gives nothing new in this case; from the latter we deduce

5. 1. LEMMA. *If the system of numbers e_{ij} is compatible with n elements of infinite order, then for any three suffixes i, j, k:*

5. 10. $$|e_{ij}|(e_{kj} \vee e_{ki}) = |e_{kj}|(e_{ij} \vee e_{ik});$$

5. 11. $$e_{ij}e_{jk}e_{ki} = e_{ji}e_{kj}e_{ik}.$$

Proof. 5. 10, written in the form $|e_{ij}|/(e_{ij} \vee e_{ik}) = |e_{kj}|/(e_{kj} \vee e_{ki})$, simply expresses the fact that the index of the three-suffix meet $(a_i) \cap (a_j) \cap (a_k)$ under the intersection $(a_i) \cap (a_k)$ may be calculated in (a_i) or in (a_k) with the same result. Thus 5. 10 holds in an amalgam.

5. 11 follows again by repeated application of the given amalgamations to the element $e_{ij}e_{jk}e_{ki}a_i$:

$$e_{ij}e_{jk}e_{ki}a_i = e_{ji}e_{jk}e_{ki}a_j = e_{ji}e_{kj}e_{ki}a_k = e_{ji}e_{kj}e_{ik}a_i;$$

thus these two multiples of a_i are the same element of the amalgam, and therefore identical, as a_i is of infinite order.

The following form of the compatibility condition derived from 5.1 will allow us to reduce the present problem to that of imbedding an amalgam of finite cycles.

We denote the greatest common divisor of e_{ij} and e_{ji} by r_{ij}, so that for all pairs i, j

5.20. $$e_{ij} = c_{ij}r_{ij}, \qquad e_{ji} = c_{ji}r_{ij}, \qquad r_{ij} = r_{ji}$$

and

5.21. $$c_{ij} \vee c_{ji} = 1.$$

5.3. LEMMA. *If the system of numbers e_{ij} is compatible with n elements of infinite order, then*

5.30 *there exist n non-zero integers $x_1, \cdots, x_n$ with greatest common divisor unity, such that $c_{ij} = x_j/(x_i \vee x_j)$ for all pairs i, j;*

5.31 *the numbers e_{ij} are also compatible with n elements of finite orders $m_i = rh_i$, where r is the least common multiple of all numbers r_{ij}, and h_i is the least common multiple of the $n-1$ numbers c_{ij} $(j = 1, \cdots, i-1, i+1, \cdots, n)$.*

Proof of 5.30. We consider the $n(n-1)/2$ linear equations

$$e_{ij}x_i - e_{ji}x_j = 0 \qquad (i, j = 1, \cdots, n; i \neq j).$$

The matrix of coefficients of this system we denote by E. E has $n(n-1)/2$ rows and n columns; the order in which the rows are taken, is of course irrelevant. Each row contains exactly two non-zero elements, e_{ij} and $-e_{ji}$; their column is given by their first suffix. We prove:

5.4. *The matrix E has rank $n-1$.*

Proof. By means of 5.11 taken for the suffixes $1, i, j$, the equation (i, j)—in obvious notation—follows from the two equations $(1, i)$ and $(1, j)$. Hence all equations follow from the $n-1$ equations $e_{1i}x_1 - e_{i1}x_i = 0$ $(i = 2, \cdots, n)$, and these are clearly independent.

Thus the equations have essentially one non-trivial solution, and necessarily all $x_i \neq 0$ by 5.0. Now let $x_1, \cdots, x_n$ denote the solution consisting of integers without common divisor (unique but for a common change of sign). Since with $e_{ij}x_i - e_{ji}x_j = 0$ also $c_{ij}x_i - c_{ji}x_j = 0$, these numbers x_i satisfy the requirements of 5.30 because of 5.21.

Proof of 5.31. We bring this proof in three steps.

1. By 4.5, it is sufficient to prove that the numbers e_{ij} are compatible with n elements of orders $p^{\lambda(m_i)}$ for any prime p. From now on let p be an arbitrary fixed prime. Put

$$\lambda(e_{ij}) = \epsilon_{ij}, \qquad \lambda(m_i) = \mu_i, \qquad \operatorname{Max} \mu_i = \mu.$$

Applying Lemmas 4.3 and 4.4 we see that we need only prove that the conditions 4.31 and 4.32 are satisfied by the numbers $\delta_{ij} = \epsilon_{ij} + \mu - \mu_i$. The third condition, 4.33, is then trivially satisfied since we even have $e_{ij}e_{jk}e_{ki} = e_{ji}e_{kj}e_{ik}$.

2. We first calculate the numbers ϵ_{ij} and δ_{ij}. Let

5.50. $$\lambda(x_i) = \alpha_i \qquad (i = 1, \cdots, n;\ x_i \text{ defined by } 5.30).$$

We assume without loss of generality

5.51. $$\alpha_1 = 0 \leq \alpha_2 \leq \cdots \leq \alpha_n.$$

5.52. Hence $\lambda(c_{ij}) = \alpha_j - \alpha_i$, if $i < j$; $\lambda(c_{ij}) = 0$, if $i > j$.

Put $\lambda(r_{ij}) = \rho_{ij}$ and $\operatorname{Max} \rho_{ij} = \rho$.

Then we obtain for the p-contents of the numbers e_{ij}:

5.53. $$\epsilon_{ij} = \alpha_j - \alpha_i + \rho_{ij}, \text{ if } i < j;\ \epsilon_{ij} = \rho_{ij}, \text{ if } i > j;$$

thus for $\mu_i = \lambda(m_i)$ from the definition of m_i, and $\mu = \operatorname{Max} \mu_i$:

5.54. $$\mu_i = \operatorname{Max}_j(\alpha_j - \alpha_i) + \rho = \alpha_n - \alpha_i + \rho,$$

and $\mu = \operatorname{Max}_i \mu_i = \alpha_n - \alpha_1 + \rho = \alpha_n + \rho$.

Hence finally:

5.55. $$\delta_{ij} = \alpha_j + \rho_{ij}, \text{ if } i < j, \text{ and } \delta_{ij} = \alpha_i + \rho_{ij}, \text{ if } i > j.$$

Thus $\delta_{ij} = \delta_{ji}$ follows immediately.

3. To prove 4.32 we need information on the numbers ρ_{ij}. We assume without loss of generality that $i < j < k$. Then

5.56. $$\delta_{ij} = \alpha_j + \rho_{ij}, \qquad \delta_{jk} = \alpha_k + \rho_{jk}, \qquad \delta_{ki} = \alpha_k + \rho_{ki}.$$

To find the relation between the numbers ρ_{ij}, ρ_{jk}, ρ_{ki}, we use the relations 5.10. for the triplets (i, j, k) and (j, i, k) of suffixes. By their nature they imply the corresponding conditions for the p-contents of the numbers e_{ij}, viz.:

$$\epsilon_{ij} + \mathrm{Min}(\epsilon_{kj}, \epsilon_{ki}) = \epsilon_{kj} + \mathrm{Min}(\epsilon_{ij}, \epsilon_{ik})$$

and

$$\epsilon_{ji} + \mathrm{Min}(\epsilon_{kj}, \epsilon_{ki}) = \epsilon_{ki} + \mathrm{Min}(\epsilon_{ji}, \epsilon_{jk}).$$

To save writing, we put for the moment $\rho_{ij} = \beta$, $\rho_{jk} = \gamma$, $\rho_{ki} = \delta$.

Then we have, because of 5. 53 and $i < j < k$:

5. 57. $$\alpha_j - \alpha_i + \beta + \mathrm{Min}(\gamma, \delta) = \gamma + \mathrm{Min}(\alpha_j - \alpha_i + \beta, \alpha_k - \alpha_i + \delta)$$

and

$$\beta + \mathrm{Min}(\gamma, \delta) = \delta + \mathrm{Min}(\beta, \alpha_k - \alpha_j + \gamma);$$

hence, subtracting $\alpha_j - \alpha_i$ from both sides of the first relation,

5. 58. $\beta + \mathrm{Min}(\gamma, \delta) = \gamma + \mathrm{Min}(\beta, \alpha_k - \alpha_j + \delta) = \delta + \mathrm{Min}(\beta, \alpha_k - \alpha_j + \gamma)$.

We show

5. 59. If $\gamma = \delta$, then $\beta \leqq \alpha_k - \alpha_j + \delta$;

if $\gamma < \delta$, then $\beta = \alpha_k - \alpha_j + \delta$;

if $\gamma > \delta$, then $\beta = \alpha_k - \alpha_j + \gamma$.

Proof. The first possibility is obvious. Otherwise, let $\gamma < \delta$; then the first of the three expressions 5. 58 equals $\beta + \gamma$; for the last to have this value, $\alpha_k - \alpha_j + \gamma \leqq \beta$ is necessary and then $\alpha_k - \alpha_j + \gamma + \delta = \beta + \gamma$ follows, and satisfies 5. 58. The last possibility follows by interchange of γ and δ: 5. 58 is symmetrical in γ and δ.

But the relations 5. 59 for the numbers ρ_{ij}, ρ_{jk}, ρ_{ki} together with 5. 56 and $\alpha_i \leqq \alpha_j \leqq \alpha_k$ prove 4. 32 for δ_{ij}, δ_{jk}, and δ_{ki}.

Now we obtain easily:

5. 6. THEOREM. *An irreducible amalgam of a finite number of infinite cycles is imbeddable in an abelian group.*

Proof. If the relations $e_{ij}a_i = e_{ji}a_j$ define an amalgam, the numbers e_{ij} satisfy the conditions of 5. 3. By 5. 32 and 4. 5, there exists an abelian group $\mathfrak{B}$ generated by n elements b_i, such that in $\mathfrak{B}$ the intersection of (b_i) and (b_j) is generated by the element $e_{ij}b_i = e_{ji}b_j$. Since $r_{ij} = e_{ij} \vee e_{ji}$ divides the orders of both b_i and b_j, the element $e_{ij}b_i - c_{ji}b_j$ of $\mathfrak{B}$ has the exact order r_{ij}.

We now form the direct sum $\mathfrak{A}$ of $\mathfrak{B}$ with one infinite cycle (a). Let $x_1, \cdots, x_n$ be integers as defined in 5. 30; the elements $a_i = x_i a + b_i$ of $\mathfrak{A}$ $(i = 1, \cdots, n)$ are of infinite order. Also because of the infinite component $x_i a$ of a_i, the intersection $(a_i) \cap (a_j)$ is generated by a multiple $rc_{ij}a_i = rc_{ji}a_j$

of $c_{ij}a_i$ and $c_{ji}a_j$. Because of the properties of b_i and b_j in $\mathfrak{B}$, $r = r_{ij}$. Thus the intersection of (a_i) and (a_j) is generated by the element $e_{ij}a_i = e_{ji}a_j$. This completes the proof.

We add a remark on the nature of the generalized free sum in the case considered in this paragraph:

5.7. THEOREM. *The generalized free sum of n infinite cycles, no two of which have zero meet, is the direct sum of one infinite cycle and at most* $n-1$ *finite cycles. It is cyclical if, and only if, the amalgamations* $e_{ij}a_i = e_{ji}a_j$ *satisfy* $e_{ij} \vee e_{ji} = 1$ *for all pairs* i, j.

Proof. Consider again the matrix E defined as the matrix of coefficients of the system of linear equations $e_{ij}x_i - e_{ji}x_j = 0$. This is clearly the characteristic matrix associated with the abelian group generated by $a_1, \cdots, a_n$ with relations $e_{ij}a_i = e_{ji}a_j$, i. e. with the generalized free sum in question. It was shown to have rank $n-1$ (cf. 5.4) which proves the first part of the theorem.

Also, if $e_{ij} \vee e_{ji} = r_{ij} \neq 1$, for some pair i, j, then the element $c_{ij}a_i - c_{ji}a_j$ is of order $r_{ij} \neq 1$, and the generalized free sum is certainly not an infinite cycle.

To prove that the group is cyclical if $e_{ij} \vee e_{ji} = 1$, i. e. $e_{ij} = c_{ij}$ in the notation used in Lemma 5.3, it is sufficient to show that the greatest common divisor of all determinants of order $n-1$ which can be formed from E is unity; or again:

If p is any prime, there is at least one determinant of order $n-1$ in E which is not divisible by p.

By 5.30 the non-zero elements of E are of the form $c_{ij} = x_j/(x_i \vee x_j)$. We put again $\lambda(x_i) = \alpha_i$, and may assume $\alpha_1 = 0 \leq \alpha_2 \leq \cdots \leq \alpha_n$. Now consider the $n-1$ rows of E containing the elements c_{2j} and $-c_{j2}$ ($j = 1, 3, \cdots, n$; the first suffix indicates the column). One of the non-zero determinants which can be formed from these rows equals, but for the sign,

$$c_{21}c_{32}\cdots c_{n2} = \frac{x_1}{x_1 \vee x_2} \cdot \frac{x_2^{n-2}}{(x_2 \vee x_3)\cdots(x_2 \vee x_n)}.$$

As p is prime to x_1, and the p-content of $x_2 \vee x_j$ ($j \geq 3$) is the same as that of x_2, this determinant is prime to p.

6. Amalgams of locally cyclic groups. Let A be an amalgam of groups $\mathfrak{G}_\alpha$ and ϕ_α any set of subgroups of $\mathfrak{G}_\alpha$. By virtue of being subgroups of a group, the groups of the set ϕ_α naturally form an amalgam. Also, all the groups of all the sets ϕ_a form an amalgam by virtue of the given amalgamations between the groups $\mathfrak{G}_a$ in the amalgam A.[4]

If, in particular, we take as ϕ_a the set of all cyclical subgroups of $\mathfrak{G}_a$, we thus obtain the amalgam of all cyclical subgroups of A. We denote it by A^*. It follows from our previous results that A^* is imbeddable in an abelian group.[5] In general, of course, this does not imply imbeddability of A: A and A^* consist of the same elements; also, two elements whose product is defined in A^*, also have a product in A (and then, of course, the same as in A^*) but, in general, the converse is not true.

We consider the extreme case that A and A^* are identical. This is true if, and only if, any two elements whose product is defined in A, also have a product in A^*, i. e. if, and only if, any two elements belonging to the same group $\mathfrak{G}_a$ belong to, and therefore generate, a cyclical subgroup of $\mathfrak{G}_a$. But this just means that every group $\mathfrak{G}_a$ is locally cyclic. Thus we have proved:

6. 0. LEMMA. *An amalgam of groups $\mathfrak{G}_a$ and the amalgam of all its cyclical subgroups are identical if, and only if, all groups $\mathfrak{G}_a$ are locally cyclic.*

6. 1. THEOREM. *An amalgam of locally cyclic groups is imbeddable in an abelian group.*

UNIVERSITY COLLEGE, HULL.

REFERENCES.

[1] R. Baer, "Free sums of groups and their generalizations," *American Journal of Mathematics*, vol. 71 (1949), pp. 706-742.

[2] Grace E. Bates, "Free loops and nets and their generalizations," *American Journal of Mathematics*, vol. 69 (1947), pp. 499-550.

[3] B. H. Neumann and Hanna Neumann, "A remark on generalized free products," *Journal of the London Mathematical Society*, vol. 25 (1950), pp. 202-204.

[4] Hanna Neumann, "Generalized free products with amalgamated subgroups, I," *American Journal of Mathematics*, vol. 70 (1948), pp. 590-625.

---

[4] This is a special case of the more general fact that an amalgam (suitably defined) of amalgams of groups is an amalgam of groups.

[5] If A itself is a group, this amalgam A^*, and therefore the abelian group freely generated by it, is an invariant of the group. I owe this remark to the referee.

GENERALIZED EUCLIDEAN SPACE IN TERMS OF A QUASI INNER PRODUCT.*

By LEONARD M. BLUMENTHAL.

1. Introduction. Necessary and sufficient conditions of various kinds have been obtained in order that an inner product may be defined in a normed linear space, with the norm and inner product related in the customary manner. Such conditions have been expressed as norm or distance relations by Fréchet [4], Jordan and von Neumann [5], and Aronszajn [1], while other writers have formulated the conditions in terms of orthogonality, etc., for spaces of more than two dimensions.

In a normed linear space S, three operations (subjected to eleven conditions) are assumed as primitive; namely, *addition* (on SS to S), *scalar multiplication* (on RS to S), and *norm* (on S to R, the set of real numbers). If an additional condition is placed on these operations (for example, the condition $\| x+y \|^2 + \| x-y \|^2 = 2(\| x \|^2 + \| y \|^2)$ of Jordan and von Neumann), then a fourth operation, *inner product* (on SS to R), may be defined in S, satisfying five relations connecting the inner product with the three primitive operations of the space.

We solve in this paper the inverse problem. An abstract set Σ is considered in which just *one* primitive operation, on $\Sigma\Sigma$ to R (quasi inner product), is asumed to satisfy *three* conditions. It is proved that if Σ is rich enough in elements to satisfy three existence postulates, then three additional properties of the quasi inner product are necessary and sufficient to permit *defining* in Σ addition, scalar multiplication, and norm so that the resulting space is a normed linear space with an inner product connected with the norm in the usual way. The concluding section of the paper is concerned with the congruent imbedding and vectorial application of Σ onto euclidean and Hilbert spaces, yielding, incidentally, a new set of postulates for Hilbert space (in terms of a single operation).

It is emphasized that our postulates and fundamental definitions are all expressed wholly and explicitly in terms of the quasi inner products of those and only those elements of Σ directly concerned. We shall advert to

---

* Received March 30, 1950. Presented to the American Mathematical Society, December 27, 1949.

this feature in the concluding section of the paper when its significance can be more readily made clear.

The first sections of the paper may be considered as giving a foundation of vector algebra in terms of one operation (quasi inner product). From that point of view our work is related to a brief sketch (without proofs) of a foundation for vector algebra given by Menger [6].

2. The postulates.

All of the postulates to be imposed are listed in this section, although they will be introduced later only as needed.

POSTULATES FOR A QUASI INNER PRODUCT.

To each pair of elements x, y of an abstract set Σ there is attached a real number (x, y), called a *quasi inner product,* in conformity with the following agreements:

Q_1 (Symmetry). *If* $x, y \varepsilon \Sigma$, *then* $(x, y) = (y, x)$.

Q_2 (Definiteness). *For each element* x *of* Σ, $(x, x) \geqq 0$.

Q_3 (Identification). *If* $x, y \varepsilon \Sigma$ *and* $(x, x) = (x, y) = (y, y)$, *then* $x = y$.

(To avoid misunderstanding, it is pointed out that Q_3 is *not* a definition of equality of elements of Σ which, as a set, is already supplied with a criterion for equality of its elements. As its listing indicates, Q_3 is an *assumption* connecting the set-equality of elements x, y with the *three* numbers (x, x), (x, y), (y, y) attached to the two elements as quasi inner products. On the other hand, we shall show later that Q_3 does, in the presence of Q_1, Q_2, S_2, E_1 (see below) actually define an equivalence relation in Σ.)

The set Σ may then be referred to as a quasi inner product space.

SCHWARZ POSTULATES.

If $x_1, x_2, \cdots, x_n \varepsilon \Sigma$, denote by $G(x_1, x_2, \cdots, x_n)$ the Gram determinant $|(x_i, x_j)|$, $(i, j = 1, 2, \cdots, n)$.

S_1. *If* $x_1, x_2 \varepsilon \Sigma$, *then* $G(x_1, x_2) \geqq 0$.

S_2. *If* $x_1, x_2, x_3 \varepsilon \Sigma$ *and* $G(x_1, x_2) = 0$, *then* $G(x_1, x_2, x_3) \geqq 0$.

S_3. *If* $x_1, x_2, x_3, x_4 \varepsilon \Sigma$ *and* $G(x_1, x_2, x_3) = 0$, *then* $G(x_1, x_2, x_3, x_4) \geqq 0$.

It is observed that S_2 and S_3 are *conditional* Schwarz inequalities, while S_1 applies to every pair of elements of Σ. The existence postulates that follow merely insure that Σ has enough elements for our purpose.

EXISTENCE POSTULATES.

If $x_1, x_2, \cdots, x_n \varepsilon \Sigma$, denote by $B(x_1, x_2, \cdots, x_n)$ the symmetric determinant obtained by bordering $G(x_1, x_2, \cdots, x_n)$ with a row and column of 1's with intersection element 0.

E_1. *There exists at least one element θ of Σ such that $(\theta, x) = 0$ for each element x of Σ.*

E_2. *For each element x of Σ and each real number λ there exists at least one element y of Σ such that $G(x, y) = 0$ and $(x, y) = \lambda \cdot (x, x)$.*

E_3. *If $x, z \varepsilon \Sigma$ and $G(x, z) \neq 0$, there exists at least one element y of Σ such that $B(x, y, z) = G(x, y, z) = 0$ and $G(x, y) = G(y, z)$.*

If scalar multiplication and addition are defined in a set, we recall that a real function $((x, y))$ defined over the set of all pairs of elements is a (real) *inner product* provided: (a) $((x, y)) = ((y, x))$, (b) $((x, x)) \geqq 0$, (c) $((x, x)) = 0$ if and only if x is the element of the space that is neutral with respect to addition, (d) $((\lambda \cdot x, y)) = \lambda \cdot ((x, y))$, and (e) $((x + y, z)) = ((x, z)) + ((y, z))$.

The reader is referred to [2] for the definition of a normed linear space.

DEFINITION 2.1. *A normed linear space in which an inner product is defined so that for each element x, the norm $\| x \|$ and inner product are connected by the relation $\| x \| = ((x, x))^{\frac{1}{2}}$, is called a generalized euclidean space.*

3. Development of the system $\{\Sigma: Q_1, Q_2, Q_3, S_2, E_1\}$.

We assume throughout this section that Σ is subjected to just the five postulates listed in the heading.

THEOREM 3.1. *There is exactly one element θ of Σ with $(\theta, x) = 0$ for each element x of Σ.*

Proof. If θ, θ' are two such elements then $(\theta, \theta) = (\theta, \theta') = (\theta'. \theta') = 0$, and so $\theta = \theta'$ by Q_3. The theorem now follows from E_1.

THEOREM 3.2. *The quasi inner product (x, x) vanishes if and only if $x = \theta$.*

Proof. By E_1, $(\theta, \theta) = 0$, and if $(x, x) = 0$ then $(x, x) = (\theta, x) = (\theta, \theta)$, which implies $x = \theta$ by Q_3.

THEOREM 3.3. *Corresponding to each element x of Σ $(x \neq \theta)$, and each real number λ, there is at most one element y of Σ such that $(x, y) = \lambda \cdot (x, x)$ and $G(x, y) = 0$.*

Proof. If $y, y^* \varepsilon \Sigma$ satisfying those two relations, then from the second one we get $(y, y) = (y^*, y^*) = \lambda^2 \cdot (x, x)$. By computation

$$G(x, y, y^*) = - (x, x) \cdot [\lambda^2 \cdot (x, x) - (y, y^*)]^2 \leqq 0,$$

by Q_2 and Theorem 3. 2. Use of S_2 gives $(y, y^*) = \lambda^2 \cdot (x, x)$, and hence $(y, y) = (y, y^*) = (y^*, y^*)$, from which $y = y^*$ follows by Q_3.

Remark. Q_3 defines an equivalence relation in $\{\Sigma : Q_1, Q_2, S_2, E_1\}$. For if $x = y$ and $y = z$ by Q_3, then $x = z$ is established by applying S_2 to $G(x, y, z)$ and using Theorems 3. 1, 3. 2 in case $(x, x) = 0$. The symmetric and reflexive properties are obvious.

DEFINITION 3. 1. *If $x \varepsilon \Sigma$, $(x \neq \theta)$, and λ is a real number, an element y of Σ is called a scalar multiple of x by λ provided $(x, y) = \lambda \cdot (x, x)$ and $G(x, y) = 0$. We write $y = \lambda \cdot x$. If $x = \theta$, put $\lambda \cdot x = \theta$, for every λ.*

THEOREM 3. 4. *For each x and each real λ, Σ contains at most one element $\lambda \cdot x$.*

Remark. For each element x *of* Σ, $0 \cdot x = \theta$.

THEOREM 3. 5. *If x, $\lambda \cdot x$, $y \varepsilon \Sigma$, $(\lambda \cdot x, y) = \lambda \cdot (x, y)$.*

Proof. The theorem being obvious if $x = \theta$, suppose the contrary. Now

$$G(x, \lambda \cdot x, y) = - (x, x) \cdot [(\lambda \cdot x, y) - \lambda \cdot (x, y)]^2 \leqq 0,$$

and using S_2 gives $(\lambda \cdot x, y) = \lambda \cdot (x, y)$.

THEOREM 3. 6. *If x, $1 \cdot x \varepsilon \Sigma$, then $1 \cdot x = x$.*

Proof. Use of Theorem 3. 5 gives $(1 \cdot x, 1 \cdot x) = (1 \cdot x, x) = (x, x)$ and $x = 1 \cdot x$ by Q_3.

THEOREM 3. 7. *If x, $\lambda_2 \cdot x$, $\lambda_1 \cdot (\lambda_2 \cdot x)$, $(\lambda_1\lambda_2) \cdot x \varepsilon \Sigma (\lambda_1, \lambda_2$ real), then $\lambda_1 \cdot (\lambda_2 \cdot x) = (\lambda_1\lambda_2) \cdot x$.*

Proof. The proof follows immediately from Q_2, Theorem 3. 5, and Q_3.

THEOREM 3. 8. *If x, $y \varepsilon \Sigma$ $(x \neq \theta)$, then y is a scalar multiple of x if and only if $G(x, y) = 0$.*

Proof. If $y = \lambda \cdot x$, then $G(x, y) = 0$ by Definition 3. 1. If $G(x, y) = 0$, put $(x, y)/(x, x) = \lambda$. Then $(x, y) = \lambda \cdot (x, x)$ which, together with $G(x, y) = 0$, gives $y = \lambda \cdot x$.

THEOREM 3.9. *If* $x, y, z \in \Sigma$, $(y \neq \theta)$, *and* $G(x,y) = G(y,z) = 0$, *then* $G(x,z) = 0$.

Proof. The theorem is trivial if $x = \theta$. In the contrary case, $y = \lambda_1 \cdot x$ and $z = \lambda_2 \cdot y$ by Theorem 3.8. Then Theorem 3.7 gives $z = \lambda_2(\lambda_1 \cdot x) = (\lambda_1 \lambda_2) \cdot x$, and $G(x,z) = 0$.

4. Introduction of a distance in $\{\Sigma: Q_1, Q_2, Q_3, S_1, S_2, E_1\}$. Attach to each two elements x, y of Σ the number

$$(*) \qquad xy = [(x,x) + (y,y) - 2(x,y)]^{\frac{1}{2}}$$

as distance. Clearly $xy = yx$ and $xx = 0$. Now

$$(\dagger) \quad (x,x) + (y,y) - 2(x,y) = [(x,x)^{\frac{1}{2}} - (y,y)^{\frac{1}{2}}]^2 + 2[(x,x)^{\frac{1}{2}}(y,y)^{\frac{1}{2}} - (x,y)] \geqq 0,$$

by S_1, and so xy is real and non-negative. Moreover, if $xy = 0$ then each summand in the right-hand member of $(\dagger)$ vanishes. This gives $(x,x) = (x,y) = (y,y)$ and consequently $x = y$. Thus the distance xy defined by $(*)$ makes Σ a *semimetric* space.

We note that $\theta x = (x,x)^{\frac{1}{2}}$ and $(x,y) = \frac{1}{2}(\theta x^2 + \theta y^2 - xy^2)$.

DEFINITION 4.1. *A distance space is metric about one of its elements* t *provided* t, u, v *are congruent with three points of the euclidean plane for every pair of elements* u, v *of the space.*

THEOREM 4.1. *The space* Σ *is metric about* θ.

Proof. Since Σ is semimetric, it suffices to show that the determinant $D(\theta, x, y)$ is negative or zero for every $x, y \in \Sigma$. (See [3, p. 56] for the definition of the Cayley-Menger determinant $D(p_1, p_2, \cdots, p_k)$ of a k-tuple of elements $p_1, p_2, \cdots, p_k$ of a semimetric space).

An easy computation shows that $D(\theta, x, y) = -4G(x,y)$, and hence $D(\theta, x, y) \leqq 0$ by S_1.

Remark. The space Σ is not necessarily metric. For if Σ consists of the quadruple θ, x_1, x_2, x_3 with $(\theta, \theta) = 0$, $(\theta, x_1) = (x_i, \theta) = 0$, $(x_i, x_i) = 16$, $(i = 1, 2, 3)$, and $(x_1, x_2) = (x_2, x_1) = (x_1, x_3) = (x_3, x_1) = (31)/2$, $(x_2, x_3) = (x_3, x_2) = (23)/2$, then Postulates $Q_1, Q_2, Q_3, S_1, S_2, E_1$ are satisfied. But $x_1x_2 = x_1x_3 = 1$, $x_2x_3 = 3$, and consequently the space is not metric.

DEFINITION 4.2. *Three elements* x, y, z *of* Σ *are algebraically linear provided* $B(x,y,z) = G(x,y,z) = 0$.

Remark. Since $B(x, y, z) = \frac{1}{4} \cdot D(x, y, z)$, elements x, y, z are *metrically* linear (that is, the sum of two of the distances xy, yz, xz equals the third) if and only if $B(x, y, z) = 0$. This, however, does not imply that $G(x, y, z) = 0$, and so algebraic linearity implies metric linearity, but not conversely.

5. Algebraic middle elements in $\{\boldsymbol{\Sigma}: Q_1, Q_2, Q_3, S_1, S_2, E_1\}$.

DEFINITION 5.1. *If $x, z \in \Sigma$ $(x \neq z)$, an element y of Σ is an algebraic middle element of x and z provided*

(a). *$G(x, z) \neq 0$, x, y, z are algebraically linear and $G(x, y) = G(y, z)$,*

or

(b). *$G(x, z) = 0$ and $y = [(1 + \lambda)/2] \cdot x$,*

where $z = \lambda \cdot x$ (with the roles of x and z reversed in the two immediately preceding equalities in case $x = \theta$).

THEOREM 5.1. *If y is an algebraic middle element of x, z then y is a metric middle element of x and z.*

Proof. From $D(x, y, z) = 4B(x, y, z) = 0$, it follows that x, y, z are metrically linear. It remains to show that $xy = yz$.

Case 1. $G(x, z) \neq 0$. Since $G(x, y, z) = 0$, the easily established relation $D(\theta, x, y, z) = 8 \cdot G(x, y, z)$ gives $D(\theta, x, y, z) = 0$, with θ, x, y, z a metric quadruple (due to Theorem 4.1 and the metric linearity of x, y, z). Hence points θ', x', y', z' of the euclidean plane exist such that the congruence

$$(*) \qquad \theta, x, y, z \approx \theta', x', y', z',$$

holds and x', y', z' lie on a straight line [3, p. 56]. Now

$$G(x, y) = -\tfrac{1}{4}D(\theta, x, y) = -\tfrac{1}{4}D(\theta', x', y') = 4\Delta^2(\theta', x', y')$$

and, similarly, $G(y, z) = 4\Delta^2(\theta', y', z')$, where $\Delta(\quad)$ denotes the area of the triangle whose vertices appear within the parentheses.

Then $\Delta(\theta', x', y') = \Delta(\theta', y', z')$, which follows from $G(x, y) = G(y, z)$, gives (using the linearity of x', y', z') $x'y' = y'z'$ and so $xy = yz$ from congruence (*).

Case 2. $G(x, z) = 0$. If $x \neq \theta$, then $y = [(1 + \lambda)/2] \cdot x$ and

$$xy = \{x, x) + [(1 + \lambda)/2]^2 \cdot (x, x) - (\lambda + 1) \cdot (x, x)\}^{\frac{1}{2}};$$

that is, $xy = \frac{1}{2} \mid 1 - \lambda \mid \cdot (x, x)^{\frac{1}{2}}$. A similar computation shows that yz has the same value. If, finally, $x = \theta$, then $y = (1/2) \cdot z$ and we get $xy = \theta y = (1/2) \cdot (z, z)^{\frac{1}{2}} = yz$.

Though the determinant D formed for a finite subset of Σ is evidently a congruence invariant, the Gram determinants G do not have this property. If, however, $x_1, x_2, \cdots, x_n$ and $y_1, y_2, \cdots, y_n$ are two n-tuples of Σ such that $x_1, x_2, \cdots, x_n \approx y_1, y_2, \cdots, y_n$ and $\theta x_i = \theta y_i$, $(i = 1, 2, \cdots, n)$, then $G(x_1, x_2, \cdots, x_n) = G(y_1, y_2, \cdots, y_n)$, since each is equal to

$$(-1)^{n+1} \cdot 2^{-n} D(\theta, x_1, x_2, \cdots, x_n).$$

Lemma 5.1. *If y is an algebraic middle element of x and z then*

$$\theta y = (1/2)[2(\theta x^2 + \theta z^2) - xz^2]^{\frac{1}{2}}.$$

Proof. Developing $D(\theta, x, y, z) = 8G(x, y, z) = 0$, and using $xy = yz = (1/2)xz$ (Theorem 5.1) the expression for θy given in the lemma is obtained.

6. The system $\{\Sigma: Q_1, Q_2, Q_3, S_1, S_2, S_3, E_1\}$. Adjoining the third Schwarz postulate to the preceding system, we prove the following theorem.

Theorem 6.1. *There is at most one algebraic middle element of two elements of Σ.*

Proof. If y, y^* are algebraic middle elements of x and z, Theorem 5.1 and Lemma 5.1 give

$$xy = yz = (1/2) \cdot xz = xy^* = y^*z, \quad \theta y = \theta y^*.$$

Case 1. $G(x, z) \neq 0$. From

$$D(\theta, x, y, z, y^*) = -16G(x, y, z, y^*) \leqq 0$$

we readily obtain $-D(\theta, x, z)(yy^*)^2 \leqq 0$. But $G(x, z) \neq 0$ gives $D(\theta, x, z) = -4G(x, z) < 0$, and it follows that $yy^* = 0$. Since Σ is semimetric, this implies $y = y^*$, and the theorem is proved in this case.

Case 2. $G(x, z) = 0$. If $x \neq \theta$ then $y = [(1 + \lambda)/2] \cdot x$, where $z = \lambda \cdot x$, and the uniqueness of y follows from Theorem 3.4. A similar argument is employed in case $x = \theta$ (and consequently $z \neq \theta$), and the proof is complete.

Definition 6.1. *Let x, z be elements of Σ with algebraic middle element y. An element $2 \cdot y$ is called a sum of x and z (written $2 \cdot y = x + z$). If $x = z$, put $x + z = 2 \cdot x$.*

If $2\cdot y = x + z$ then $(1/2)(2\cdot y) = (1/2)(x+z)$ or (Theorem 3.7) $y = \frac{1}{2}(x+z)$.

THEOREM 6.2. *Two elements x, z of Σ have at most one sum $x + z$.*

LEMMA 6.1. *If x, y, $x + y$ are elements of Σ, then*

$$(x+y, x) = (x, x) + (x, y), \quad (x+y, x+y) = (x, x) + (y, y) + 2(x, y).$$

Proof. Writing $2\cdot w = x + y$, we have

$$(x+y, x) = (2\cdot w, x) = 2\cdot(w, x) = \theta w^2 + \theta x^2 - xw^2.$$

Since $xw^2 = (1/4)xy^2$ and (Lemma 5.1) $\theta w^2 = \frac{1}{2}(\theta x^2 + \theta y^2) - (1/4)\cdot xy^2$, substitution yields

$$(x+y, x) = \theta x^2 + (1/2)(\theta x^2 + \theta y^2 - xy^2) = (x, x) + (x, y).$$

Also,

$$\begin{aligned}(x+y, x+y) &= (2\cdot w, 2\cdot w) = 4\cdot(w, w) = 4\cdot\theta w^2 = 2(\theta x^2 + \theta y^2) - xy^2 \\ &= 2(\theta x^2 + \theta y^2) - (\theta x^2 + \theta y^2 - 2(x, y)) = \theta x^2 + \theta y^2 + 2\cdot(x, y) \\ &= (x, x) + (y, y) + 2\cdot(x, y).\end{aligned}$$

THEOREM 6.2. *If $x, y, x + y, z \varepsilon \Sigma$, then $(x+y, z) = (x, z) + (y, z)$.*

Proof. Suppose, first, that $G(x, y) \neq 0$. Since

$$G(x, y, x+y) = 4\cdot G(x, y, \tfrac{1}{2}(x+y)) = 0,$$

S_3 yields $G(x, y, x+y, z) \geqq 0$.

Use of Lemma 6.1 gives

$$G(x, y, x+y, z) = -[(x+y, z) - (x, z) - (y, z)]^2\cdot G(x, y) \leqq 0.$$

Consequently $G(x, y, x+y, z) = 0$ and $(x+y, z) = (x, z) + (y, z)$.

If $G(x, z) = 0$, then (assuming $x \neq \theta$), $\frac{1}{2}(x+y) = [(1+\lambda)/2]\cdot x$ with $y = \lambda\cdot x$, and so $x + y = (1+\lambda)\cdot x$ (Theorem 3.7). Whence

$$\begin{aligned}(x+y, z) &= ((1+\lambda)\cdot x, z) = (1+\lambda)\cdot(x, z) \\ &= (x, z) + (\lambda\cdot x, z) = (x, z) + (y, z).\end{aligned}$$

The desired result is obtained in a similar manner in case $x = \theta$ (and consequently $y \neq \theta$).

7. The system $\{\Sigma: Q_1, Q_2, Q_3, S_1, S_2, S_3, E_1, E_2, E_3\}$. We adjoin now the two existence postulates E_2 and E_3 to our system.

THEOREM 7. 1. *For each element x of Σ and each real number λ, Σ contains exactly one element $\lambda \cdot x$. If $x, y \in \Sigma$ there is exactly one element $x + y$ of Σ.*

Proof. The proof of the first statement of the theorem is an immediate consequence of Postulate E_2, Definition 3. 1, and Theorem 3. 4, while the proof of the second statement follows from the first one and Theorem 6. 1, when it is shown that at least one algebraic middle element exists for every pair of distinct elements of Σ. (If $x = z$, $x + z = 2 \cdot x$ exists by the first part of the theorem).

If $x, z \in \Sigma$ $(x \neq z)$, Postulate E_3 insures the existence of at least one middle element of x, z whenever $G(x, z) \neq 0$. If x, $z \in \Sigma$, $(x \neq z)$ with $G(x, z) = 0$, then $z = \lambda \cdot x$ (assuming $x \neq \theta$) and the element $[(1 + \lambda)/2] \cdot x$ which exists by the first part of the theorem, is the desired algebraic middle element of x and z. If, finally, $x = \theta$ then the element $(1/2) \cdot z$ is an algebraic middle element of x and z.

THEOREM 7. 2. *Addition and scalar multiplication in Σ have the following properties:*

$$\text{(i)} \qquad x + y = y + x,$$

$$\text{(ii)} \qquad x + (y + z) = (x + y) + z,$$

$$\text{(iii)} \qquad x + y = x + z \textit{ implies } y = z,$$

$$\text{(iv)} \qquad \lambda \cdot (x + y) = \lambda \cdot x + \lambda \cdot y,$$

$$\text{(v)} \qquad (\lambda_1 + \lambda_2) \cdot x = \lambda_1 \cdot x + \lambda_2 \cdot x.$$

Proof. Properties (i), (ii), (iv), (v) follow at once from Postulates Q_1, Q_3, and Theorems 3. 5, 6. 2. To establish (iii), we have $x + y = x + z$ implies $(x + y, y) = (x + z, y)$, and consequently $(y, y) = (y, z)$. Also, $(x + y, z) = (x + z, z)$ and so $(y, z) = (z, z)$. Hence $y = z$ by Q_3.

THEOREM 7. 3. *Distances θx, $x \in \Sigma$, have the following properties:*

$$\text{(a)} \qquad \theta x = (x, x)^{\frac{1}{2}} \geqq 0,$$

$$\text{(b)} \qquad \theta x = 0 \textit{ if and only if } x = \theta,$$

(c) $$\text{dist.}\,[\theta, \lambda \cdot x] = |\lambda| \cdot \theta x,$$

(d) $$\text{dist.}\,[\theta, x + y] \leqq \theta x + \theta y.$$

Proof. The relation $\theta x = (x, x)^{\frac{1}{2}}$ is given in the second paragraph of Section 4, and (b) follows from Q_2 and Theorem 3. 2.

Since

$$\begin{aligned}\text{dist.}\,[\theta, \lambda \cdot x] &= [(\theta, \theta) + (\lambda \cdot x, \lambda \cdot x) - 2(\theta, \lambda \cdot x)]^{\frac{1}{2}} \\ &= (\lambda \cdot x, \lambda \cdot x)^{\frac{1}{2}} = |\lambda| \cdot (x, x)^{\frac{1}{2}} = |\lambda| \cdot \theta x,\end{aligned}$$

property (c) is proved.

To establish (d) we have from $G(x, y) \geqq 0$,

$$[(x, x)^{\frac{1}{2}} + (y, y)^{\frac{1}{2}}]^2 \geqq (x, x) + (y, y) + 2\,|(x, y)|,$$

or

(*) $$\theta x + \theta y \geqq (x + y, x + y)^{\frac{1}{2}} = \text{dist.}\,[\theta, x + y].$$

We are justified by Theorem 7. 3 in defining the *norm* $\| x \|$ of an element x of Σ as the distance θx. Noting that

$$\begin{aligned}(x - y, x - y)^{\frac{1}{2}} &= (x + (-1)y, x + (-1)y)^{\frac{1}{2}} \\ &= [(x, x) + (y, y) - 2(x, y)]^{\frac{1}{2}} = xy,\end{aligned}$$

or $xy = \| y - x \| = \text{dist.}\,[\theta, y - x]$, it follows that for each three elements x, y, z of Σ

$$\text{dist.}\,[y - x, z - x] = \| (z - x) - (y - x) \| = \| z - y \| = yz.$$

Remark. The space Σ is metric.

THEOREM 7. 4. *The space* $\{\Sigma\colon Q_1, Q_2, Q_3, S_1, S_2, S_3, E_1, E_2, E_3\}$ *is a generalized euclidean space.*

Proof. Postulates Q_1, Q_2 and Theorems 3. 2, 3. 5, 6. 2 prove that the quasi inner product (x, y) has the properties (a)-(e) of an inner product (Section 2), and Theorems 3. 6, 3. 7, 7. 2) establish that Σ is a linear space with respect to the addition $x + y$ and the scalar multiplication $\lambda \cdot x$.

By Theorem 7. 3 the space Σ is normed by taking $\| x \| = \theta x$, and since $\theta x = (x, x)^{\frac{1}{2}}$, Σ satisfies all the requirements of Definition 2. 1 for a generalized euclidean space.

8. Some further properties of Σ. We understand now by Σ a set subjected to all nine of the postulates in Section 2; that is, Σ is a generalized euclidean space (Theorem 7.4). It is well-known that if such a space is separable and complete, it is congruent with a euclidean space or with Hilbert space according as its dimension is finite or infinite, respectively. It might be of interest in this concluding section to give another proof of this important theorem on the basis of the theory we have been developing.

THEOREM 8.1. *Metric linearity and algebraic linearity are equivalent in* Σ.

Proof. By the Remark following Definition 4.2, algebraic linearity implies metric linearity. To prove the converse it suffices to show that if $x, y, z \in \Sigma$, then $B(x, y, z) = 0$ implies $G(x, y, z) = 0$. This is obvious in case the elements are not pairwise distinct, or if one of the elements is θ. Suppose neither alternative holds.

By the concluding remarks of Section 7, it is easily seen that $x, y, z \approx \theta, y - x, z - x$, and so the metric linearity of x, y, z implies that of $\theta, y - x, z - x$. Consequently $G(y - x, z - x) = -(1/4)D(\theta, y - x, z - x) = 0$, and application of Theorem 3.8 gives

$$z - x = \lambda \cdot (y - x); \quad z = (1 - \lambda) \cdot x + \lambda \cdot y,$$

from which $G(x, y, z) = 0$ follows by computation.

DEFINITION 8.1. *A space has the weak euclidean four-point property provided each quadruple of its points containing a linear triple is congruent with four points of the euclidean plane.*

THEOREM 8.2. *The space* Σ *has the weak euclidean four-point property.*

Proof. Let x, y, z, w be any four points of Σ (pair-wise distinct) with x, y, z metrically linear. Since $x, y, z, w \approx x - w, y - w, z - w, \theta$, it suffices to prove the latter quadruple imbeddable congruently in the plane. The space Σ being metric, this follows upon showing $D(\theta, x - w, y - w, z - w) = 0$, [3].

Now the metric linearity of x, y, z implies that of $x - w$, $y - w$, $z - w$, and hence (Theorem 8.1) this triple is algebraically linear; that is $G(x - w, y - w, z - w) = 0$. But

$$D(\theta, x - w, y - w, z - w) = 8 \cdot G(x - w, y - w, z - w) = 0,$$

and the proof is complete.

Since for each two of its distinct elements x, z the space Σ contains an algebraic (and hence metric) middle element, Σ is *metrically convex.* If, moreover, $x, y \in \Sigma$ the element $z = 2y - x$ is such that y is the middle element of x and z, and so Σ is *externally convex.*

DEFINITION 8.2. *An element x of Σ is a limit of an infinite sequence $\{x_n\}$ of elements of Σ provided* $\lim(x, x_n) = \lim(x_n, x_n) = (x, x)$.

Remark. Writing $(x, x) + (x_n, x_n) - 2(x, x_n)$ as in (†) of Section 4, it is seen that $\lim x_n = x$ if and only if $\lim xx_n = 0$.

Cauchy sequences and separability are now easily defined in terms of inner product, and we have the following theorems.

THEOREM 8.3. *The space Σ is congruent with Hilbert space provided it is separable, complete, and for each positive integer k a subset $x_1, x_2, \cdots, x_k$ of Σ exists with* $G(x_1, x_2, \cdots, x_k) \neq 0$.

THEOREM 8.4. *The space Σ is congruent with the n-dimensional euclidean space E_n provided it is complete, and n is the smallest positive integer such that each $n+1$ elements of Σ have a vanishing determinant G.*

Since Σ is metrically convex, externally convex, and has the weak euclidean four-point property, these theorems are immediate consequences of metric characterization theorems of Hilbert and euclidean spaces (with respect to the class of all semimetric spaces), respectively [3, p. 70], observing (as noted in Section 5) that for each positive integer k,

$$D(\theta, x_1, x_2, \cdots, x_k) = (-1)^{k+1} \cdot 2^k G(x_1, x_2, \cdots, x_k).$$

It is observed that a congruence of Σ with E_n or with Hilbert space preserves inner products if and only if the element θ' corresponding to θ in the congruence is the origin (or null vector) of the space. Since this can always be brought about by a translation, Theorems 8.3 and 8.4 can be restated in terms of a *vectorial application* of the vector space Σ onto the euclidean or Hilbert vector spaces.

In conclusion, we wish to point again to the fact that we have avoided the obvious and ad hoc procedure that, for example, defines an element c as a sum of elements a, b provided $(c, x) = (a, x) + (b, x)$ *for every element x of Σ.* Such a device places a condition on the inner products of a, b, c with *every* element of the space in order to ascertain if one of the three elements is a sum of the other two! This question is answered, according to the procedure of the paper, merely by examining the inner products attached to a, b, and c, the three elements concerned. The linearity of inner product clearly partakes more of the nature of a theorem than a postulate, and much of the paper is concerned with establishing that property from our (a priori) milder assumptions.

THE UNIVERSITY OF MISSOURI.

REFERENCES.

[1] N. Aronszajn, "Caractérisation métrique de l'espace de Hilbert, des espaces vectoriels et de certains groupes métriques," *Comptes Rendus, Academie des Sciences*, vol. 201 (1935), pp. 811-813; pp. 873-875.

[2] S. Banach, *Théorie des operations linéaires*, Warsaw, 1932.

[3] L. M. Blumenthal, *Distance Geometries*, University of Missouri Studies, vol. 13, 1938.

[4] M. Fréchet, "Sur la définition axiomatique d'une classe d'espaces vectoriels distanciés applicables vectoriellement sur l'espace de Hilbert," *Annals of Mathematics*, vol. 36 (1935), pp. 705-718.

[5] P. Jordan and J. von Neumann, "On inner products in linear metric spaces," *ibid.*, vol. 36 (1935), pp. 719-723.

[6] K. Menger, "Neuer Aufbau der Vektoralgebra," *Ergebnisse eines mathematischen Kolloquiums* (Wien), Heft 5 (1933), pp. 27-29; *L'Enseignement Mathématique*, nos. 5-6 (1936), pp. 359-362.

ON THE SPECTRUM OF HILBERT'S MATRIX.*[1]

By Wilhelm Magnus.

---

1. Introduction and summary. It has been shown by Hilbert that the infinite matrix

$$(1.1)\qquad (\alpha_{nm}),\ \text{where}\ \alpha_{nm} = (n+m+1)^{-1}; \qquad n, m = 0, 1, 2, \cdots,$$

is bounded, and Hilbert's inequality,

$$(1.2)\qquad 0 \leqq \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} x_n x_m/(n+m+1) \leqq \pi \sum_{n=0}^{\infty} x^2{}_n,$$

for real x_n, has been proved in many different ways; for references cf. Hardy, Littlewood and Pólya [2]. The finite segments of (1.1) have been investigated by H. Frazer [1] and O. Taussky [6]. The matrix (1.1) is associated with Legendre's polynomials (cf. Szegö [4]), and the quadratic form in (1.2) has been used by Szegö [5] for a normalization of Hankel forms.

Hilbert's original proof of (1.2) (cf. H. Weyl [7], I. Schur [3]) suggests that the spectrum of (1.1) is purely continuous. We shall obtain the following result:[2]

Theorem 1. *The spectrum of Hilbert's matrix* (1.1) *is purely continuous. Every real value of* λ *for which* $0 \leqq \lambda \leqq \pi$ *belongs to the spectrum.*

2. An integral equation. We shall prove

Theorem 2. *If the real numbers* g_n, h_n, $(n = 0, 1, 2, \cdots)$ *satisfy*

$$(2.1)\qquad \sum_{0}^{\infty} g^2{}_n = M < \infty, \qquad \sum_{0}^{\infty} h^2{}_n = N < \infty$$

---

* Received January 25, 1950.

[1] Work sponsored by the Office of Naval Research.

[2] I am indebted to Professor Wintner for calling my attention to the fact that the second statement of Theorem 1 can be derived from Carleman's work (*Sur les équations intégrales singulières à noyau réel et symmétrique*, Uppsala, 1923, pp. 169) on the integral analogue of Hilbert's matrix. Carleman shows that the kernel $(s+t)^{-1}$, $0 \leqq s$, $t < \infty$, has the purely continuous spectrum $[0, \pi]$; from this, from Hilbert's original proof of (1.2), and from a result due to H. Weyl (*Rendiconti del Circolo Matematico di Palermo*, vol. 27 (1909), pp. 373-392), it follows that every point of $[0, \pi]$ belongs to the spectrum of (1.1).

(*which implies that the power series*

$$g(z)=\sum_0^\infty g_n z^n, \qquad h(z)=\sum_0^\infty h_n z^n \tag{2.2}$$

converge for $|z|<1$), *and if* λ *denotes a real number such that* $0\leqq\lambda\leqq\pi$, *then neither of the integral equations*

$$\int_0^1 g(t)/(1-tz)\,dt=\lambda g(z), \tag{2.3}$$

$$\int_0^1 h(t)/(1-tz)\,dt=\lambda h(z)+1 \tag{2.4}$$

has a solution, except for the solution $g(z)\equiv 0$ *of* (2.3).

This is equivalent to Theorem 1. We can show this by expanding $(1-tz)^{-1}$ in a series of ascending powers of zt, introducing this expansion into (2.3) and (2.4), and observing that, because of (2.1), term by term integration is permitted.

To prove Theorem 2, we deduce first from (2.1) and from Schwarz's inequality that, for $|z|<1$,

$$|g(z)|\leqq M^{\frac{1}{2}}|1-z^2|^{-\frac{1}{2}}, \qquad |h(z)|\leqq N^{\frac{1}{2}}|1-z^2|^{-\frac{1}{2}}. \tag{2.5}$$

This shows that we can iterate (2.3) and (2.4) which gives

$$\begin{aligned}\int_0^1 \{g(t)/(t-z)\}\log\{(1-z)/(1-t)\}\,dt&\\ =\lambda^2 g(z),&\end{aligned} \tag{2.6}$$

$$\begin{aligned}\int_0^1 \{h(t)/(t-z)\}\log\{(1-z)/(1-t)\}\,dt&\\ =\lambda^2 h(z)+\lambda-z^{-1}\log(1-z).&\end{aligned} \tag{2.7}$$

We now introduce

$$1-t=e^{-\tau}, \qquad 1-z=e^{-\zeta}, \qquad g(t)=\gamma(\tau), \qquad h(t)=\eta(\tau). \tag{2.8}$$

Multiplying (2.6) and (2.7) by $e^{-p\zeta}$ and integrating with respect to ζ from 0 to ∞, we find

$$\begin{aligned}&\lambda^2\int_0^\infty \gamma(\zeta)e^{-p\zeta}d\zeta\\ &=\int_0^\infty \gamma(\tau)e^{-p\tau}d\tau\{\int_0^\infty e^{-p\sigma}\sigma/(1-e^{-\sigma})\,d\sigma\\ &\qquad+\int_0^\infty e^{-\sigma(1-p)}\sigma d\sigma/(1-e^{-\sigma})-\int_\tau^\infty e^{-\sigma(1-p)}\sigma/(1-e^{-\sigma})\,d\sigma\}.\end{aligned} \tag{2.9}$$

The exchanging of the integrations is permitted if $\operatorname{Re} p > \frac{1}{2}$, where Re denotes the real part.

The integrals from 0 to ∞ with respect to σ in (2. 9) can be evaluated explicitly, and the integral from τ to ∞ can be evaluated by an expansion of $\{1 - \exp(-\sigma)\}^{-1}$ into a sum of terms $\exp(-n\sigma)$, provided that $\operatorname{Re} p < 1$. The result is

$$\{(\pi/\sin \pi p)^2 - \lambda^2\} \int_0^\infty e^{-p\zeta}\gamma(\zeta)\,d\zeta = \sum_{k=0}^{\infty} \{b_k(1-p+k)^{-1} + a_k(1-p+k)^{-2}\}, \tag{2.10}$$

where $\frac{1}{2} < \operatorname{Re} p < 1$ and

$$a_k = \int_0^\infty \gamma(\sigma)e^{-(k+1)\sigma}d\sigma = \int_0^1 g(t)(1-t)^k dt, \tag{2.11}$$

$$b_k = \int_0^\infty \gamma(\sigma)e^{-(k+1)\sigma}d\sigma = -\int_0^1 g(t)(1-t)^k \log(1-t)\,dt. \tag{2.12}$$

But since the integral on the left-hand side of (2. 9) is an analytic function of p if $\operatorname{Re} p > \frac{1}{2}$, the latter condition is the only restriction for the validity of (2. 10), the right hand side being regular in p except for poles at $p = k + 1$ $(k = 0, 1, 2, \cdots)$. Since the left-hand side in (2. 10) vanishes at

$$p = \tfrac{1}{2} + n \pm i\theta, \qquad (\theta = \pi^{-1}\{\log(\pi + (\pi^2 - \lambda^2)^{\frac{1}{2}}) - \log \lambda\}), \tag{2.13}$$

where $n = 1, 2, \cdots$, the same must hold for the right-hand side of (2. 10).

We can deduce from (2. 1) that an application of the inversion formula of the Laplace transform

$$\int_0^\infty e^{-p\zeta}\gamma(\zeta)\,d\zeta \tag{2.14}$$

must give $\gamma(\zeta)$. If we compute (2. 14) from (2. 10), this supplies a representation of $\gamma(\zeta)$ by an integral taken over a parallel to the imaginary axis of the p-plane, which then can be evaluted by an application of the formula of residues. To prove this, we deduce from (2. 5) that

$$|a_k| \leqq M^{\frac{1}{2}}(k + \tfrac{1}{2})^{-1}, \qquad |b_k| \leqq M^{\frac{1}{2}}(k + \tfrac{1}{2})^{-2}, \tag{2.15}$$

which shows that the series on the right-hand side of (2. 10) tends to zero if $\operatorname{Re} p < 1$ and $|p| \to \infty$. Going back from ζ to z, the result is, for $0 < \lambda < \pi$,

$$g(z) = \tfrac{1}{2}\lambda^{-2}(\pi^2 - \lambda^2)^{-\frac{1}{2}} \sum_{m=0}^{\infty} (1-z)^{m-\frac{1}{2}}\{q_m(1-z)^{-i\theta} + \bar{q}_m(1-z)^{i\theta}\}, \tag{2.16}$$

where θ is defined by (2.13) and q_n by

$$(2.17)\qquad q_m = -i\sum_{k=0}^{\infty}\{a_k(\tfrac{1}{2}+m+k-i\theta)^{-2}+b_k(\tfrac{1}{2}+m+k-i\theta)^{-1}\}.$$

In a similar way, we find from (2.7) that

$$(2.18)\qquad h(z) = \tfrac{1}{2}\lambda^{-2}(\pi^2-\lambda^2)^{-\frac{1}{2}}\sum_{m=0}^{\infty}(1-z)^{m-\frac{1}{2}}\{r_m(1-z)^{-i\theta}+\bar{r}_m(1-z)^{i\theta}\},$$

where

$$(2.19)\qquad r_m+\lambda(\tfrac{1}{2}-m+i\theta)^{-1}+i\sum_{k=0}^{\infty}(\tfrac{1}{2}+k-m+i\theta)^{-2}$$
$$= -i\sum_{k=0}^{\infty}\{\alpha_k(\tfrac{1}{2}+m+k-i\theta)^{-2}+\beta_k(\tfrac{1}{2}+m+k-\theta)^{-1}\},$$

$$(2.20)\qquad \alpha_k = \int_0^1 h(t)(1-t)^k dt, \qquad \beta_k = -\int_0^1 h(t)(1-t)^k \log(1-t)dt.$$

Finally, it follows from (2.3), by Stieltjes' inversion formula, that, if $w > 1$,

$$(2.21)\qquad \lim_{\epsilon\to 0}\{g(w+i\epsilon)-g(w-i\epsilon)\} = 2\pi i(w/\lambda)g(w^{-1});$$

where $g(w \pm i\epsilon)$ denotes the analytic continuation of $g(z)$ into the neighborhood of the upper and lower side of the real axis. (2.21) is also satisfied if we substitute $h(z)$ for $g(z)$. Combined with (2.16), this gives

$$(2.22)\qquad w\sum_{m=0}^{\infty}(-1)^m(w-1)^{m-\frac{1}{2}}\{q_m(w-1)^{-i\theta}+\bar{q}_m(w-1)^{i\theta}\}$$
$$= \sum_{m=0}^{\infty}(1-w^{-1})^{m-\frac{1}{2}}\{q_m(1-w^{-1})^{-i\theta}+\bar{q}_m(1-w^{-1})^{i\theta}\}.$$

From this we see, by letting $w \to 1$, that if $q_0 = 0$, then $q_1 = 0$. But q_0 must be zero, because

$$(2.33)\qquad \int_0^1\{g(z)\}^2dz = \sum_{n,m=0}^{\infty}g_n g_m(n+m+1)^{-1} \leqq \pi\sum_0^{\infty}g^2{}_n$$

is finite, according to (1.2) and (2.1). On the other hand, it follows from (2.16) that the integral on the left in (2.23) is infinite unless $q_0 = 0$. Hence, if $g(z)$ satisfies (2.1) and (2.3), and if $0 < \lambda < \pi$, then

$$(2.24)\qquad g(1) = g'(1) = 0, \qquad |g'(z)| \leqq M',$$

where M' is a constant and $0 \leqq z \leqq 1$. Similarly, we find

$$(2.25)\qquad h(1) = h'(1) = 0, \qquad |h'(z)| \leqq N'.$$

Differentiating (2.3) and (2.4) with respect to z, we have from (2.24), (2.25)

$$(2.26)\quad \lambda g'(z) = \int_0^1 (1-tz)^{-2}\{tg(t)\}dt$$
$$= -z^{-1}\int_0^1 (1-zt)^{-1}\{tg'(t)+g(t)\}dt,$$

and (2.26) is also valid if we substitute $h(z)$ for $g(z)$.

If we combine (2.26) and (2.3), we obtain

$$\lambda\{zg'(z)+g(z)\} = \int_0^1 tg'(t)(1-tz)^{-1}dt.$$

Since $|g'(t)| \leqq M'$, this gives $\lambda(n+1)|g_n| \leqq M'\int_0^1 t^{n+1}dt$, which shows that

$$(2.27)\qquad \sum_0^\infty \{(n+1)g_n\}^2 < \infty.$$

We can replace the second equation in (2.26) by an infinite system of linear equations, viz.,

$$\sum_{m=0}^\infty (m+1)g_m(n+m+1)^{-1} = -\lambda n g_n, \qquad (n=0,1,2,\cdots).$$

Putting $(m+1)g_m = x_m$, we find

$$(2.28)\qquad \sum_{n=0}^\infty\sum_{m=0}^\infty x_n x_m/(n+m+1) = -\lambda\sum_{n=0}^\infty n/(n+1)x_n^2.$$

Because of (2.27), we can apply (1.2), obtaining a contradiction unless $x_1 = x_2 = \cdots = 0$. Hence we see from (2.28) that also $x_0 = 0$, and therefore $g(z) \equiv 0$. The proof for $h(z) \equiv 0$ (i.e., for the non-existence of $h(z)$) is precisely the same. This proves Theorem 2 for $0 < \lambda < \pi$.

To complete the proof of Theorem 2, it will be sufficient to show that $g(z) \equiv 0$ if $\lambda = 0$ or $\lambda = \pi$, since the spectrum is a closed set. Therefore we have to exclude only the possibility that $\lambda = 0$ or $\lambda = \pi$ belongs to the point spectrum of (1.1). The case $\lambda = 0$ can easily be excluded by introducing $G(z) = \sum_0^\infty g_n z^{n+1}(n+1)^{-1}$. This is a continuous function of z for $0 \leqq z \leqq 1$, and from (2.3) we have in the case $\lambda = 0$

$$\int_0^1 g(t)t^n dt = G(1) - n\int_0^1 G(t)t^{n-1}dt = 0$$

for $n = 0, 1, 2, \cdots$. For $n = 0$, this gives $G(1) = 0$, and for $n = 1, 2, \cdots$ we find that all the moments of $G(t)$ vanish, i. e., $G(t) \equiv 0$. The case $\lambda = \pi$ can be dealt with in the same way as the general case if we modify (2.16) according to the fact that the evaluation of $\gamma(\zeta)$ from (2.10) requires the computation of residues at poles of the second order of the integrand. The analogue of (2.16) then involes a term with $\log(1 - z)$.

California Institute of Technology.

REFERENCES.

[1] H. Frazer, "Note on Hilbert's inequality," *Journal of the London Mathematical Society,* vol. 21 (1946), pp. 7-9.

[2] G. H. Hardy, J. E. Littlewood, G. Pólya, *Inequalities,* Cambridge (1934), pp. 226-227.

[3] I. Schur, "Bemerkungen zur Theorie der beschränkten Bilinearformen mit unendlich vielen Veränderlichen," *Journal für die reine und angewandte Mathematik,* vol. 140 (1911), pp. 1-28.

[4] G. Szegö, *Orthogonal polynomials,* New York (1939).

[5] ———, "A Hankel-Féle Formákról," *Mathematikai es Természettudományi értesitö,* vol. 36 (1918), pp. 497-538. Review in *Jahrbuch über die Fortschritte der Mathematik,* vol. 46 (1916-1918), p. 649.

[6] O. Taussky, "A remark concerning the characteristic roots of the finite segments of the Hilbert matrix," *Quarterly Journal of Mathematics,* Oxford ser., vol. 20 (1949), pp. 80-83.

[7] H. Weyl, *Singuläre Integralgleichungen mit besonderer Berücksichtigung des Fourierschen Integraltheorems.* Dissertation, Göttingen, 1908, pp. 83-86.

THE EXPANSION THEOREM FOR SIGMA-MONOGENIC FUNCTIONS.*

By Lipman Bers.

---

1. Introduction. A Σ-monogenic function $f(z)$ is a complex-valued function $f = u + iv$ whose real and imaginary part satisfy the partial differential equations

$$(1.1) \qquad \sigma_1(x)u_x = \tau_1(y)v_y, \qquad \sigma_2(x)u_y = -\tau_2(y)v_x,$$

[1, 2, 3].[1] In this note we assume that the functions $\sigma_1, \tau_1; \sigma_2, \tau_2$ are of class C'' and positive, and we say that f is Σ-monogenic only if u and v are of class C''. Without essential loss of generality we may assume that the functions σ, τ are defined for all real values of their arguments. If u and v are of class C'' and satisfy instead of (1.1) the system

$$(1.1') \qquad u_x/\sigma_2(x) = \tau_1(y)v_y, \qquad u_y/\sigma_1(x) = -\tau_2(y)v_x,$$

we say that $f = u + iv$ is Σ'-monogenic.[2]

The Σ-integral of a continuous function $f = u + iv$ is defined by the relation

$$(1.2) \qquad \int f d_\Sigma z = \int u\sigma_2 dx - v\tau_2 dy + i\int (v/\sigma_1)dx + (u/\tau_1)dy.$$

It is path-independent if $f(z)$ is Σ-monogenic and is in this case a Σ'-monogenic function of the upper (variable) limit of integration. The Σ-derivative[3] of a Σ-monogenic function $f(z)$ is defined by

$$(1.3) \qquad d_\Sigma f/d_\Sigma z = f^{[1]}(z) = \sigma_1 u_x + iv_x/\sigma_2 = \tau_1 v_y - iu_y/\tau_2.$$

The real and imaginary parts of $f^{[1]}$ satisfy (1.1'), so that if they are of class C'', $f^{[1]}$ is Σ'-monogenic. The higher Σ, Σ'-derivatives of the Σ-monogenic function $f(z)$, provided they exist, are defined by the recurrence formulae $f^{[0]} = f$, $f^{[2n+1]} = d_\Sigma f^{[2n]}/d_\Sigma z$, $f^{[2n]} = d_{\Sigma'} f^{[2n-1]}/d_{\Sigma'} z$.

A significant class of Σ-monogenic and Σ'-monogenic functions are the

---

* Received January 6, 1950; revised March 30, 1950. Presented to the Society on August 30, 1949. The research leading to this paper was supported by the ONR (Contract N6-onr-248).

[1] Numbers in brackets refer to the references at the end of the paper.

[2] The letters Σ, Σ' denote the coefficient-arrays of the systems (1.1), (1.1'). Note that $\Sigma'' = \Sigma$.

[3] Σ-integration and Σ-differentiation are inverse processes.

"formal powers" defined (for any constant values of a and z_0) by the recurrence formulae[4]

$$a \cdot Z^{(0)}(z_0; z) = a \cdot \tilde{Z}^{(0)}(z_0; z) = a,$$

$$a \cdot Z^{(n)}(z_0; z) = n \int_{z_0}^{z} a \cdot \tilde{Z}^{(n-1)}(z_0; \zeta)\, d_{\Sigma'}\zeta,$$

$$a \cdot \tilde{Z}^{(n)}(z_0; z) = n \int_{z_0}^{z} a \cdot Z^{(n-1)}(z_0; \zeta)\, d_{\Sigma}\zeta.$$

The following facts are immediate consequences of this definition. For a real λ and complex a, b

$$\lambda\{a \cdot Z^{(n)}(z_0; z)\} = (\lambda a) \cdot Z^{(n)}(z_0; z), \tag{1.4}$$

$$a \cdot Z^{(n)}(z_0; z) + b \cdot Z^{(n)}(z_0; z) = (a + b) \cdot Z^{(n)}(z_0; z). \tag{1.5}$$

A "formal polynomial"

$$P(z) = \sum_{\nu=0}^{N} a_\nu \cdot Z^{(\nu)}(z_0; z) \tag{1.6}$$

is a Σ-monogenic function possessing Σ, Σ'-derivatives of all orders. In particular,

$$P^{[\nu]}(z_0) = \nu! a_\nu, \qquad \nu = 0, 1, \cdots, N;\ P^{[\mu]}(z) \equiv 0,\ \mu > N. \tag{1.7}$$

The formal powers satisfy the inequality

$$|\, a \cdot Z^{(n)}(z_0; z) \,| \leqq |\, a \,| \{C(|\, z_0 \,| + |\, z \,|)\,|\, z - z_0 \,|\}^n, \tag{1.8}$$

where $C(r)$ is a positive continuous function depending upon the functions σ, τ [1, p. 76]. This implies that if $|\, a_\nu \,|^{1/\nu} = O(1)$, $\nu \to \infty$, then the "formal power series"

$$f(z) = \sum_{\nu=0}^{\infty} a_\nu \cdot Z^{(\nu)}(z_0; z) \tag{1.9}$$

converges uniformly and absolutely in some neighborhood G of z_0, and so do all series obtained from it by Σ, Σ'-differentiation. Hence $f(z)$ is Σ-monogenic in G, possesses Σ, Σ'-derivatives of all orders, and

$$f^{[\nu]}(z_0) = \nu! a_\nu, \qquad \nu = 0, 1, \cdots. \tag{1.10}$$

Now let $f(z)$ be a given Σ-monogenic function defined in the neighborhood of z_0. Can it be expanded in a formal power series of the form (1.9)? An affirmative answer was given (by A. Gelbart and the author) under the hypothesis that the coefficients σ, τ are *analytic* functions [**1**, p. 83]. In this note we prove the same expansion theorem under the much weaker assumption that the coefficients σ_1, σ_2, τ_1, τ_2 are twice continuously differentiable.

---

[4] Note that $a \cdot Z^{(n)}$ is one symbol. In general, $aZ^{(n)} \neq a \cdot Z^{(n)}$.

2. The expansion theorem. Consequences. Our aim is to prove the following theorem.

EXPANSION THEOREM. *Let $f(z)$ be Σ-monogenic in a domain D. In the neighborhood of every point z_0 of D, $f(z)$ admits a unique (uniformly and absolutely convergent) expansion of the form* (1.9). *If $|f(z)| < M$ in D, then*

$$|a_\nu| < A^\nu M, \tag{2.1}$$

where A is a constant depending only on D, the functions σ_i, τ_j, and the distance from z_0 to the boundary of D.

With the aid of this theorem many results obtained previously for the case of analytic σ, τ can be reëstablished without the analyticity hypothesis. This applies in particular to the analogue of Morera's theorem [**1**, Theorem 3.6] and to the theorem on the quasi-analytic character of Σ-monogenic functions [**1**, p. 84]. The inequality (2.1) in conjunction with (1.10) implies the analogue of Weierstrass' convergence theorem.

The expansion theorem implies that a Σ-monogenic function possesses Σ, Σ' derivatives of *all* orders. Noting that the real part of a Σ-monogenic function is a solution of the equation

$$\{u_x\sigma_1(x)/\tau_1(y)\}_x + \{u_y\sigma_2(x)/\tau_2(y)\}_y = 0, \tag{2.2}$$

We obtain very precise information about the way in which differentiability properties of the coefficients of the elliptic equation (2.2) are reflected in the differentiability properties of solutions. We mention explicitly only an extreme case.

For

$$\sigma_1 = \sigma_2 = 1, \quad \tau_1 = \tau_2 = \exp\int\alpha(y)\,dy, \tag{2.3}$$

equation (2.2) becomes

$$\Delta u = \alpha(y)u_y. \tag{2.4}$$

On the other hand (2.3) implies that $a \cdot Z^{(n)}(z_0, x + iy_0) = a(x - x_0)^n$. Thus we obtain the following theorem: *if $\alpha(y)$ is of class C', every solution of* (2.4) *is analytic in x.*

3. Lemmas. In this section we shall prove several lemmas which are needed for the proof of the expansion theorem.

LEMMA 1. *Let* D_0 *be a bounded domain,* D *a subdomain of* D_0, z_0 *a point of* D, r_0 *the distance from* z_0 *to the boundary of* D. *Let* $p(x,y)$ *be a bounded continuous function defined in* D_0, P *a bound for* $|p(x,y)|$, $\phi(x,y)$ *a bounded function of class* C'' *defined in* D *and satisfying the equation.*

$$\Delta\phi = p\phi. \tag{3.1}$$

There exists a constant A *depending only on* D_0 *and* P *such that the inequality*

$$|\phi(x,y)| \leqq M \text{ in } D, \tag{3.2}$$

implies that

$$|\phi_x(x_0,y_0)|,\ |\phi_y(x_0,y_0)| \leqq AM/r_0. \tag{3.3}$$

Proof.[5] Set $\rho(x,y;\xi,\eta) = (4\pi)^{-1}\log[(x-\xi)^2 + (y-\eta)^2]$ and

$$\omega(x,y) = \int\int_D \rho(x,y;\xi,\eta)p(\xi,\eta)\phi(\xi,\eta)\,d\xi d\eta.$$

There exists a constant B depending only upon D_0 such that in D

$$|\omega(x,y)|,\ |\omega_x(x,y)|,\ |\omega_y(x,y)| \leqq BPM. \tag{3.4}$$

We have, for instance,

$$|\omega(x,y)| \leqq PM\int\int_D |\rho(x,y;\xi,\eta)|\,d\xi d\eta \leqq PM\int\int_{D_0} |\rho(x,y;\xi,\eta)|\,d\xi d\eta,$$

and the last integral has a finite maximum on the closure of D_0.

On the other hand, $h(x,y) = \phi(x,y) - \omega(x,y)$ is harmonic in D, and since, by (3.2) and (3.4), $|h| \leqq (1+BP)M$, it follows[6] that

$$|h_x(x_0,y_0)|,\ |h_y(x_0,y_0)| \leqq 2(1+BP)M/r_0. \tag{3.5}$$

From (3.4) and (3.5) the assertion follows.

Consider now a Σ-monogenic function $f = u + iv$. The function u is a solution of (2.2). Set

$$x_1 = \int_0^x [\sigma_2(t)/\sigma_1(t)]^{\frac{1}{2}}\,dt, \qquad y_1 = \int_0^y [\tau_2(t)/\tau_1(t)]^{\frac{1}{2}}\,dt,$$

and $\alpha = [\sigma_1(x)\sigma_2(x)/\tau_1(y)\tau_2(y)]^{1/4}$, $\phi = \alpha u$. Then ϕ considered as a function of

---

[5] The lemma is hardly new. The proof is sketched since the author knows of no reference.

[6] See [5], p. 321 where the analogous inequality is proved for three dimensions.

x_1 and y_1 satisfies the equation $\partial^2\phi/\partial x_1^2 + \partial^2\phi/\partial y_1^2 = (\partial^2\alpha/\partial x_1^2 + \partial^2\alpha/\partial y_1^2)\phi/\alpha$, so that Lemmas 1 may be applied to ϕ. A similar reasoning holds for Σ'-monogenic functions. Noting the definition of Σ-derivatives, we obtain the following result:

LEMMA 2. *Let D be a domain situated within the rectangle $a \leqq x \leqq a'$, $b \leqq y \leqq b'$, z_0 a point in D, r_0 the distance from z_0 to the boundary of D. Let $f(z)$ be Σ-monogenic (or Σ'-monogenic) in D. If in D, $|\operatorname{Re} f(z)| < M$, then $|f^{[1]}(z_0)| < AM/r_0$ where A is a constant depending only on the rectangle and on the behaviour of the functions σ, τ in the intervals $[a, a']$, $[b, b']$, respectively.*[7]

The following is an immediate consequence of Lemma 2:

LEMMA 3. *If $f_n(z)$, $n = 1, 2, \cdots$, and $f(z)$ are Σ-monogenic in a domain D, if $\operatorname{Re} f_n \rightarrow \operatorname{Re} f$ uniformly in D, and if, at some point z_0 of D, $f_n(z_0) \rightarrow f(z_0)$, then $f_n \rightarrow f$ uniformly in every closed subdomain of D.*

As another application of Lemma 2 we shall prove the inequality (2.1) of the expansion theorem for the case of a formal polynomial.

LEMMA 4. *Let D be a bounded domain containing the point z_0, $P(z)$ a formal polynomial of the form (1.6), and M an upper bound for $|P(z)|$ in D. Then*

$$|a_\nu| \leqq A^\nu M, \qquad \nu = 0, 1, \cdots, N, \tag{3.6}$$

where A is a constant of the kind described in the expansion theorem.

Proof. For $\nu = 0$ (3.6) holds for every $A \geqq 1$. Now let ν be a fixed integer ($0 < \nu \leqq N$). Let $\bar{D}_\mu$ denote the disc $|z - z_0| \leqq (1 - \mu/\nu) r_0$ where r_0 is the distance from z_0 to the boundary of D, and $\mu = 0, 1, \cdots, \nu - 1$. Finally, let M_μ denote the maximum of $|P^{[\mu]}(z)|$ in $\bar{D}_\mu$, and let A' be the constant from Lemma 2. By repeated application of this lemma we obtain the inequalities $M_1 \leqq (\nu A'/r_0) M_0$, $M_2 \leqq (\nu A'/r_0) M_1 \leqq (\nu A'/r_0)^2 M_0, \cdots$, and finally $|P^{[\nu]}(z_0)| \leqq (A'/r_0)^\nu \nu^\nu M_0$. Since $M_0 \leqq M$ and a_ν is given by (1.7), we have $|a_\nu| \leqq (A'/r_0)^\nu (\nu^\nu/\nu!) M$, and since by Stirling's formula $\nu^\nu/\nu! \sim e^\nu (2\pi\nu)^{-\frac{1}{2}}$, the assertion follows.

The proof of the expansion theorem will be based on Lemma 4 and on the following lemma which establishes the possibility of expanding a Σ-monogenic function in a series of formal polynomials.

---

[7] An analysis of the argument will show that the maxima of the second derivatives of σ, τ enter in A.

LEMMA 5. *Let $f(z)$ be Σ-monogenic in the closed rectangle $\bar{R}: a \leqq x \leqq a'$ $b \leqq y \leqq b'$. Then there exists a sequence of formal polynomials $P_n(z)$ of the form* (1.6) *which converges to $f(z)$ uniformly in every closed subdomain of the open rectangle R: $a < x < a'$, $b < y < b'$.* (The point z_0 is not required to belong to $\bar{R}$).

Proof. By subtracting from $f(z)$ a formal polynomial of the fourth degree, we may obtain a function which vanishes at the vertices[8] of R. We may assume, therefore, that

$$f(a+ib) = f(a'+ib) = f(a+ib') = f(a'+ib') = 0.$$

Since the Dirichlet problem is always solvable for equation (2.2), we may represent $u = \operatorname{Re} f(z)$ in R as a sum of four solutions of (2.2), each of which vanishes on all but one side of $\bar{R}$. It will be sufficient to consider one such function. Thus we assume now that $u(x,y) = \operatorname{Re} f(z)$ satisfies the boundary conditions:

$$u(a,y) = u(a',y) = 0, \quad b \leqq y \leqq b'; \qquad u(x,b') = 0, \quad a \leqq x \leqq b',$$
$$u(x,b) = \Phi(x) \in C'', \quad a \leqq x \leqq a', \quad \Phi(a) = \Phi(a') = 0.$$

A function

$$u(x,y) = \phi(x)\psi(y) \tag{3.7}$$

will be a solution of (2.2) if

$$[\sigma_1(x)\phi'(x)]' + \lambda\sigma_2(x)\phi(x) = 0, \tag{3.8}$$

$$[\psi'(y)/\tau_2(y)]' - \lambda\psi(y)/\tau_1(y) = 0. \tag{3.9}$$

The characteristic values of (3.8) for the boundary conditions $\phi(a) = \phi(a') = 0$ are simple and positive. Let us denote them by Λ_j^2, $j = 1, 2, \cdots$, and let $\phi_j(x)$ be the corresponding (normalized) characteristic functions. Since Λ_j^2 is certainly not a characteristic value of (3.9) for the boundary conditions $\psi(b) = \psi(b') = 0$, there exists a solution $\psi_j(y)$ of (3.9) for $\lambda = \Lambda_j^2$ which satisfies the conditions $\psi_j(b) = 1$, $\psi_j(b') = 0$. It is well known that $\Phi(x)$ admits the uniformly and absolutely convergent expansion $\Phi(x) = \sum_{j=1}^{\infty} \alpha_j\phi_j(x)$, $a \leqq x \leqq a'$. The series $\sum_{j=1}^{\infty} \alpha_j\phi_j(x)\psi_j(y)$ converges to $u(x,y)$ uniformly on the boundary of R. The same must be true throughout R since solutions of (2.2) are known to obey the maximum-minimum principle.

---

[8] See the interpolation theorem in [1], p. 89. This theorem is proved without using analyticity of the functions σ, τ.

Now we determine Σ-monogenic functions $f_j(z)$ such that

$$\operatorname{Re} f_j(z) = \alpha_j \phi_j(x) \psi_j(y), \qquad j = 1, 2, \cdots, \tag{3.10}$$

$$\operatorname{Im} f_1(z_1) = \operatorname{Re} f(z_1), \qquad \operatorname{Im} f_j(z_1) = 0, \qquad j = 2, 3, \cdots, \tag{3.11}$$

where z_1 is some fixed point in R. By Lemma 3 we conclude that $f(z) = \sum_{j=1}^{\infty} f_j(z)$, the series being uniformly convergent in every closed subdomain of R.

The assertion of Lemma 5 would be proved if we could find for every positive ϵ and for every j a formal polynomial of the form (1.6) which in R differs from $f_j(z)$ by less than ϵ. That this can be done, however, follows from the fact that every Σ-monogenic function satisfying (3.10) can be expanded in a formal power series of the form (1.9) which converges uniformly in every bounded domain.[9]

4. Proof of the expansion theorem.[10]

Let $f(z)$ be Σ-monogenic in the domain D containing z_0, and let $|f(z)|$ not exceed M in D. According to Lemma 5 there exists a sequence of formal polynomials,

$$P_n(z) = \sum_{\nu=0}^{N_n} a_\nu^{(n)} \cdot Z^{(\nu)}(z_0; z),$$

such that

$$f(z) = \lim_{n \to \infty} P_n(z), \tag{4.1}$$

uniformly in the closure of some neighborhood G_0 of z_0. We set $a_\nu^{(n)} = 0$ for $\nu > N_n$. Let A' denote the constant of Lemma 4 for the domain G_0 and the point z_0. Then, by (4.1) and Lemma 4,

$$|a_\nu^{(n)}| \leqq A'^\nu (M + \epsilon_n); \qquad \lim_{n \to \infty} \epsilon_n = 0. \tag{4.2}$$

The limit

$$\lim_{n \to \infty} a_\nu^{(n)} = a_\nu \tag{4.3}$$

exists for $\nu = 0, 1, \cdots$. In fact, we have that

$$|a_\nu^{(n)} - a_\nu^{(m)}| \leqq A'^\nu \cdot \max |P_n(z) - P_m(z)| \, (z \,\varepsilon\, \bar{G}_0). \tag{4.4}$$

From (4.2) we conclude that

$$|a_\nu| \leqq A'^\nu M, \tag{4.5}$$

---

[9] See [1], Theorem 10. 2. The proof of this theorem uses no analyticity assumptions.

[10] I am indebted to Professor Hermann Weyl for suggesting a simplification of the argument.

so that the formal power series

$$(4.6)\qquad \sum_{\nu=0}^{\infty} a_\nu \cdot Z^{(\nu)}(z_0; z)$$

converges uniformly and absolutely in some neighborhood G_1 of z_0. Let r be a positive number such that the disc $G: |z - z_0| < r$ is contained in G_1 and $rA'C(2|z_0| + r) < 1$, where C is the function in inequality (1.8). For $m \to \infty$, (4.4) yields

$$|a_\nu - a_\nu^{(n)}| \leqq A'^\nu \cdot \max |P_n(z) - f(z)| = A'^\nu \eta_n, \qquad \eta_n \to 0.$$

Using this inequality, we see from (1.5) and (1.8) that, for z in G,

$$|\sum_{\nu=0}^{\infty} a_\nu \cdot Z^{(\nu)}(z_0; z) - P_n(z)| \leqq \sum_{\nu=0}^{\infty} |(a_\nu - a_\nu^{(n)}) \cdot Z^{(\nu)}(z_0; z)|$$

$\leqq \eta_n/\{1 - rA'C(2|z_0| + r)\}$, so that, by (4.1), $f(z) = \sum_{\nu=0}^{\infty} a_\nu \cdot Z^{(\nu)}(z_0; z)$. The uniqueness of this expansion follows at once from Lemma 4.

It remains to prove inequality (2.1). (Note that it is not implied by (4.5) since A' depends upon G_0). We observe first that since $f(z)$ may be expanded in a formal power series in the neighborhood of every point of D, all Σ, Σ' derivatives of f exist. Next we repeat verbatim the proof of Lemma 4, replacing P by f and using (1.10) instead of (1.7).

The proof of the expansion theorem is now complete. The method used to establish this result ought to suffice to free the expansion theorems obtained by Diaz [4] and Protter [6] for their generalizations of Σ-monogenic functions from unnecessary analyticity conditions. It remains an open question, however, whether the conditions imposed by us on the functions σ, τ can be substantially weakened.

Syracuse University.

REFERENCES.

[1] L. Bers and A. Gelbart, "On a class of functions defined by partial differential equations," *Transactions of the American Mathematical Society*, vol. 56 (1944), pp. 67-93.

[2] ———, "On a class of differential equations in mechanics of continua," *Quarterly of Applied Mathematics*, vol. 1 (1943), pp. 168-188.

[3] ———, "On generalized Laplace transformations," *Annals of Mathematics*, vol. 48 (1947), pp. 342-357.

[4] J. B. Diaz, "On a class of partial differential equations of even order," *American Journal of Mathematics*, vol. 68 (1946), pp. 611-659.

[5] O. D. Kellogg, *Foundations of Potential Theory*, Berlin, 1929.

[6] M. H. Protter, "Generalized spherical harmonics," *Transactions of the American Mathematical Society*, vol. 63 (1948), pp. 314-341.

ON THE SPECTRUM OF A BOUNDARY VALUE PROBLEM WITH TWO SINGULAR ENDPOINTS.*

By KENNETH G. WOLFSON.

1. Let a positive $p(t)$ and a real-valued $q(t)$, where $-\infty < t < \infty$, be continuous functions with the property that the differential equation

$$(px')' + (q + \lambda)x = 0 \tag{1}$$

is of the "Grenzpunkt" type for both $t = -\infty$ and $t = +\infty$. Thus (1) determines a boundary value problem in the Hilbert space $L^2(-\infty, +\infty)$; that is, $L(z) = -(pz')' - qz$ is a self-adjoint operator defined on the set of functions $z(t)$ for which z, pz' are absolutely continuous; and z, $L(z)$ are of class $L^2(-\infty, \infty)$. Let the spectrum of this operator be denoted by S and the set of its cluster points by S'. The assumptions above also imply that (1) and a fixed homogeneous boundary condition,

$$x(c) \cos \alpha + x'(c) \sin \alpha = 0, \tag{2}$$

where $-\infty < c < \infty$ and $0 \leqq \alpha < \pi$, determine two boundary value problems, one in each of the Hilbert spaces $L^2(-\infty, c)$ and $L^2(c, \infty)$. Let $S^{c\alpha}, S_{c\alpha}$ denote the spectra of the respective problems. The derived sets $S^{c\alpha\prime}$, $S_{c\alpha}{}'$ are independent of c and α ([8], p. 251) and may be denoted by S^*, S_*, respectively.

The object of this paper will be to determine the relationship between the sets S', S^*, S_* and those between the set S and the spectra $S_a{}^b$ of the Sturm-Liouville problems belonging to (1) and the boundary conditions

$$x(a) = 0 \quad \text{and} \quad x(b) = 0, \tag{3}$$

where $-\infty < a < b < \infty$.

The second part of the problem is analogous to the corresponding problem concerning $S_{c\alpha}$ treated in [2]. The results below will be dependent on those of [2].

For $\lambda_1 < \lambda_2$ and $a < b$, let $N(a, b, \lambda_1, \lambda_2)$ be the number of points of $S_a{}^b$ in the closed interval $[\lambda_1, \lambda_2]$. Put $n(\lambda_1, \lambda_2) = \liminf N(a, b, \lambda_1, \lambda_2)$, where $a \to -\infty$ and $b \to +\infty$. The following theorems will be proved:

---

* Received March 29, 1950. Presented to the American Mathematical Society, April 28, 1950.

(I) *A necessary and sufficient condition for* $\lambda = \lambda_0$ *to be in* S *is that*

$$(4) \qquad n(\lambda_0 - \epsilon, \lambda_0 + \epsilon) \geqq 1 \text{ for every } \epsilon > 0.$$

(II) *A necessary and sufficient condition for* $\lambda = \lambda_0$ *to be in* S' *is that*

$$(5) \qquad n(\lambda_0 - \epsilon, \lambda_0 + \epsilon) = \infty \text{ for every } \epsilon > 0.$$

Let $\delta(\lambda_0)$ denote the number (0, 1 or 2) of the half lines $(-\infty, 0]$, $[0, \infty)$ on which (1) is oscillatory for λ near, but exceeding, λ_0. Put $m(\lambda_1, \lambda_2) = \limsup N(a, b, \lambda_1, \lambda_2)$, where $a \to -\infty$ and $b \to \infty$. It turns out that condition (4) is equivalent to

$$(4') \qquad m(\lambda_0 - \epsilon, \lambda_0 + \epsilon) \geqq \delta(\lambda_0) + 1 \text{ for every } \epsilon > 0.$$

It may be pointed out that, in (I), the boundary conditions (3) cannot be replaced by arbitrary homogeneous boundary conditions, say

$$(6) \qquad x(a)\cos\beta + x'(a)\sin\beta = 0, \qquad x(b)\cos\beta + x'(b)\sin\beta = 0.$$

Corresponding to the differential equation $x'' + \lambda x = 0$ and the boundary conditions (6) determined by $\beta = \pi/4$, the number $\lambda = -1$ is an eigenvalue for every choice of a, b; but the set S is the half-line $0 \leqq \lambda < \infty$, and does not contain the point $\lambda = -1$.

It will be clear that the proof of (I) can be modified so as to imply the following theorem. This theorem is the analogue of one concerning S_{ca}, which was proved first in [8], pp. 252-257, with restrictions on $q(t)$ and λ_2, and without restrictions in [2], p. 917.

(III) *If* $n(\lambda_1, \lambda_2) < \infty$, *put* $N(\lambda) = n(\lambda_1, \lambda)$ *for* $\lambda_1 < \lambda < \lambda_2$. *Then* (i) $N(\lambda)$ *is a non-decreasing (integral valued) step-function*; (ii) $N(\lambda)$ *is continuous from the left*; (iii) *every jump (if any) of* $N(\lambda)$ *has the value* 1; *finally,* (iv) *the discontinuity points of* $N(\lambda)$ *on* $\lambda_1 < \lambda < \lambda_2$ *are in the point spectrum and no other* λ *values of the open interval* (λ_1, λ_2) *are in the spectrum* S *of* (1).

The following theorem will also be proved:

(IV) $S' = S_* + S^*$, *where the addition is set-theoretic.*

The set S' can be decomposed into the cluster points of the point spectrum and the continuous spectrum C. According to Weyl ([9], p. 450), each λ in C is at most of double multiplicity, as defined by Hellinger [5]. If $p \equiv 1$, while q is even and periodic, then it is known [10] that the sets S_* and S^* are identical and consist entirely of continuous spectra. Furthermore every point λ of $S_* = S^*$ is also in C with multiplicity two; cf. the spectral resolution given in [6], p. 934. Thus suggests the possibility that, in the general case, (IV) has some interpretation as regards the multiplicities of points of the continuous spectrum C. This will remain an open question.

As a consequence of (IV), results on the set S^* can be carried over to the set S'. For example, by the use of the theorem (*) of [3] we have:

If there exists a number μ such that for $\lambda < \mu$ every solution of (1) has only a finite number of zeros for $-\infty < t < \infty$, but has infinitely many zeros when $\lambda > \mu$, then μ is the least point of S'. If such a number μ fails to exist, then S clusters at $\lambda = -\infty$.

2. Proof of the necessity of (4) in (I). If $\liminf N(a, b, \lambda_1, \lambda_2) = k$, where $a \rightarrow -\infty$ and $b \rightarrow +\infty$, then the open interval (λ_1, λ_2) contains at most k points of S.

This can be established as follows: Let K_{ab} denote the bounded integral operator whose kernel on the rectangle $a \leqq s$, $t \leqq b$ is the Green function determined by (1) for $\lambda = i$ and boundary conditions (3). This operator determines the spectrum $S_a{}^b$. Let K denote the corresponding integral operator for the interval $(-\infty, \infty)$; so that $K = (L - iI)^{-1}$, where I is the identity operator. For convenience, suppose that the domain of K_{ab} is not $L^2(a, b)$ but $L^2(-\infty, +\infty)$; and that, for any element g of this Hilbert space, $K_{ab}g$ is 0 outside the interval $[a, b]$. From the form of the Green functions and from the definition of "Grenzpunkt" ([8], p. 226), it can be verified that K_{ab} converges to K, as $a \rightarrow -\infty$ and $b \rightarrow +\infty$, in the sense of strong convergence of bounded operators on the Hilbert space $L^2(-\infty, \infty)$; cf. the derivation of the spectral resolution of L as given by Titchmarsh [7], pp. 56-57. The statement at the beginning of this section follows from known properties of strongly convergent bounded operators.

This assures the necessity of (4).

3. Proof of sufficiency of (4) in (I). The proof will be divided into three parts (i), (ii), (iii) corresponding to the cases $\delta(\lambda_0) = 0, 1, 2$ respectively.

(i) If $n(\lambda_0 - \epsilon, \lambda_0 + \epsilon) \geqq 1$ for a given $\epsilon > 0$, then there exist numbers A_ϵ, B_ϵ such that, for all $a \leqq A_\epsilon$ and $b \geqq B_\epsilon$, the set $S_a{}^b$ has at least one point in the interval $[\lambda_0 - \epsilon, \lambda_0 + \epsilon]$. By Theorem (I) in [2], the set S^b, where $S^b = S^{b0}$, contains at least one point in $[\lambda_0 - \epsilon, \lambda_0 + \epsilon]$ for every $b \geqq B_\epsilon$. By the Lemma in **3** of [1], there exist solutions of (1) of class $L^2(-\infty, b)$, corresponding to the values λ lying in the interval $[\lambda_0 - \epsilon, \lambda_0 + \epsilon]$ for sufficiently small $\epsilon > 0$. Since (1) is of the *Grenzpunkt* type at $t = -\infty$, these solutions are unique up to constant factors. Denote by $x(t, \lambda)$ a solution of (1) which is $L^2(-\infty, b)$. Since $\delta(\lambda_0) = 0$, the solution $x(t, \lambda)$ has only a finite number of zeros on $-\infty < t < \infty$. The proof can be completed by adapting the method used in [1] for the analogous problem concerning the sets $S_{ca}{}^b$ and S_{ca}, for fixed c. That method is based on the following: If $x^*(t, \lambda)$ denotes a solution of (1) and (2), then the n-th largest zero of x^* on $c < t < \infty$ is a continuous decreasing function of λ, and the Wronskian of $x^*(t, \lambda^1)$ and $x^*(t, \lambda^2)$ is zero at $t = c$. In the present case of the function $x(t, \lambda)$, the point $t = -\infty$ replaces the point $t = c$. The n-th largest zero of $x(t, \lambda)$ on $-\infty < t < \infty$ is a continuous decreasing function of λ. Also, the Wronskian of $x(t, \lambda^1)$ and $x(t, \lambda^2)$ satisfies

$$(7) \qquad p(t)[x'(t, \lambda^2)x(t, \lambda^1) - x(t, \lambda^2)x'(t, \lambda^1)] \to 0 \text{ as } t \to -\infty;$$

[8], p. 242; cf. formula (19) in [4].

The details of the proof for case (i) will not be given as they follow closely those of [1].

(ii) In the case $\delta(\lambda_0) = 1$, it can be assumed that (1) is non-oscillatory on the half axis $-\infty < t \leqq 0$. Suppose, if possible, that (4) holds but that λ_0 is not in S. Then there exists an $\epsilon > 0$ such that the interval $[\lambda_0 - \epsilon, \lambda_0 + \epsilon]$ is free of points of S. Since $n(\lambda_0 - \epsilon, \lambda_0 + \epsilon) \geqq 1$, there exist numbers A_ϵ, B_ϵ with the property that, whenever $a \leqq A_\epsilon$, $b \geqq B_\epsilon$, the set $S_a{}^b$ has at least one point in $[\lambda_0 - \epsilon, \lambda_0 + \epsilon]$. Theorem (I) in [2] implies that, for every $b \geqq B_\epsilon$, the set S^b has at least one point in the interval $[\lambda_0 - \epsilon, \lambda_0 + \epsilon]$. It will be shown that there exist values of b corresponding to which S^b has at least two points in that interval. Let $\{\mu_h(b)\}$ denote the set of eigenvalues in S^b which are less than the least point of $S^{b'} = S^*$. The continuity and monotony of the n-th zero of $x(t, \lambda)$, mentioned above, and the fact that the number of zeros on $(-\infty, b)$ of eigenfunctions corresponding to successive elements of the sequence $\{\mu_h(b)\}$ differ by one, allow an adaptation of the arguments of [2], p. 916, as follows:

Let $x(t, \lambda)$ have the same meaning as in (i). Consider a value $b' > b$ satisfying $x(b', \lambda_0 - \epsilon) = 0$; so that $\lambda_0 - \epsilon = \mu_h(b')$ for some value of h. Then $\mu_{h+1}(b') \leqq \lambda_0 + \epsilon$. For if this were not the case, it would follow that $\mu_h(b'') < \lambda_0 - \epsilon$, while $\mu_{h+1}(b'') > \lambda_0 + \epsilon$ for b'' near, but exceeding, b'. This contradicts the fact that for every $b \geqq B_\epsilon$ the set S^b contains at least one element in $[\lambda_0 - \epsilon, \lambda_0 + \epsilon]$. Hence, for some arbitrarily large b, S^b has at least two points in $[\lambda_0 - \epsilon, \lambda_0 + \epsilon]$.

To complete the proof, it will be shown that the assumption that no point of S is in the interval $[\lambda_0 - \epsilon, \lambda_0 + \epsilon]$ leads to a contradiction.

Since the addition of a constant λ_0 to $q(t)$ simultaneously translates the sets S^b, S by $-\lambda_0$, it can be assumed that $\lambda_0 = 0$. Thus, for some choice of b and h, the numbers $\mu_h(b)$, $\mu_{h+1}(b)$ lie in the interval $[-\epsilon, \epsilon]$, but this interval contains no points of S. Let $x_h(t)$, $x_{h+1}(t)$ be eigenfunctions corresponding to μ_h, μ_{h+1} respectively. Put

$$(8) \qquad z(t) = c_1 x_h(t) + c_2 x_{h+1}(t) \text{ or } z(t) = 0$$

according as $-\infty < t \leqq b$ or $b < t < \infty$, where c_1, c_2 are chosen so that $z'(b-0) = 0$ (and $z(t) \not\equiv 0$). Then z, pz' are absolutely continuous, and z, $L(z)$ are of class $L^2(-\infty, \infty)$. Hence $z(t)$ belongs to the domain of the self-adjoint operator $L(z)$. From (1) and (8)

$$L(z) = c_1 \mu_h x_h(t) + c_2 \mu_{h+1} x_{h+1}(t) \text{ or } L(z) = 0,$$

according as $-\infty < t \leqq b$ or $b < t < \infty$. Since $|\mu_h|$, $|\mu_{h+1}| \leqq \epsilon$, it follows from (8) that

$$(9) \qquad \int_{-\infty}^{+\infty} L^2(z)\,dt \leqq \epsilon^2 \int_{-\infty}^{+\infty} z^2 dt,$$

since x_h, x_{h+1} are orthogonal on $(-\infty, b)$. Since $[-\epsilon, \epsilon]$ is free of points of S, the spectral resolution of the operator $L(z)$ implies

$$(10) \qquad \int_{-\infty}^{+\infty} L^2(z)\,dt > \epsilon^2 \int_{-\infty}^{+\infty} z^2 dt.$$

This inequality contradicts (9) and completes the proof of the case (ii).

(iii) Assume that $\delta(\lambda_0) = 2$, that (4) holds and that, if possible, λ_0 is not in S. Then there is an interval $[\lambda_0 - \epsilon, \lambda_0 + \epsilon]$ free of points of S; and there exist numbers A_ϵ, B_ϵ such that for all $a \leqq A_\epsilon$ and $b \geqq B_\epsilon$, the interval $[\lambda_0 - \epsilon, \lambda_0 + \epsilon]$ contains at least one point of $S_a{}^b$. The theorem (I) in [2] implies that $[\lambda_0 - \epsilon, \lambda_0 + \epsilon]$ contains at least one point $\lambda = \lambda_1$ of S^b, and at least one point $\lambda = \lambda_2$ of S_a. Let $x_1 = x_1(t)$, $x_2 = x_2(t)$ denote corresponding eigenfunctions. Let $d < a$ be a zero of x_1 and $e > b$ be a zero of x_2. These zeros exist in view of the assumption $\delta(\lambda_0) = 2$. Let $x_3 = x_3(t)$ denote an eigenfunction corresponding to a point $\lambda = \lambda_3$ of $S_d{}^e$ lying in the interval $[\lambda_0 - \epsilon, \lambda_0 + \epsilon]$. Finally, let $z(t) = c_1 x_1(t), x_3(t), c_2 x_2(t)$ for $-\infty < t \leqq d$, $d < t \leqq e$, and $e < t < \infty$, respectively, where c_1, c_2 are chosen so as to make $z'(t)$ continuous. Since x_1, x_2, x_3 satisfy (1) for $\lambda = \lambda_1, \lambda_2, \lambda_3$, respectively, it follows that $(pz')' = 0$ for $t = d, e$. Hence $(pz')'$ is continuous; so that z, pz' are absolutely continuous on $-\infty < t < \infty$. Thus

$z(t)$ is in the domain of the operator $L(z)$. By a procedure analogous to that used in (ii), one obtains a contradiction to the fact that $[\lambda_0 - \epsilon, \lambda_0 + \epsilon]$ is free of points of S. This completes the proof of case (iii).

4. Proof of (II). The necessity of (5) follows from **2**. The proof of (I) can be modified so as to imply that if $n(\lambda_1, \lambda_2) = k$, then there are k points of S in $[\lambda_1, \lambda_2]$; cf. (III). This shows the sufficiency of (5).

5. Proof of (IV). Assume that λ_0 is in S_*. Theorem (I) of [2] implies that if c is fixed, $-\infty < c < \infty$, then there exist corresponding sets $S_c{}^b$ having arbitrarily many points in $[\lambda_0 - \epsilon, \lambda_0 + \epsilon]$ for every $\epsilon > 0$. Hence $n(\lambda_0 - \epsilon, \lambda_0 + \epsilon) = \infty$ for every $\epsilon > 0$, and so λ_0 is in S' by (II). Similarly, if λ_0 is in S^*, it is in S'.

Assume that λ_0 is in S'. It remains to verify that λ_0 is in either S^* or S_*. Suppose, if possible, that this is not the case. If c is a fixed number, there exists an interval $[\lambda_0 - \epsilon, \lambda_0 + \epsilon]$ which contains at most one point of each of the sets S^c, S_c. Then no set $S_a{}^c$ or $S_c{}^b$ for any $a < c < b$ can have more than two points in the interval $[\lambda_0 - \epsilon, \lambda_0 + \epsilon]$. But then Sturm's comparison theorem shows that $N(a, b, \lambda_0 - \epsilon, \lambda_0 + \epsilon) \leqq 5$ for all $a < c < b$, which is contradictory to the fact that λ_0 is assumed to be in S', and so $n(\lambda_0 - \epsilon, \lambda_0 + \epsilon) = \infty$.

6. Proof of equivalence of conditions (4) and (4′). The proof will again be divided into three parts (i), (ii), (iii) corresponding to $\delta(\lambda_0) = 0, 1, 2$ respectively.

(i) In this case it is only necessary to show that if, for some $\epsilon > 0$, the limit superior $m(\lambda_0 - \epsilon, \lambda_0 + \epsilon) \geqq 1$, then $n(\lambda_0 - \epsilon, \lambda_0 + \epsilon) \geqq 1$, for the converse is obvious. The assumption $m(\lambda_0 - \epsilon, \lambda_0 + \epsilon) \geqq 1$ implies that there exists a sequence of intervals $(a_1, b_1), (a_2, b_2), \cdots$ such that $a_k > a_{k+1} \to -\infty$ and $b_k < b_{k+1} \to \infty$, as $k \to \infty$, and that $S_a{}^b$, where $(a, b) = (a_k, b_k)$, has at least one point $\lambda = \lambda_k$ in the closed interval $[\lambda_0 - \epsilon, \lambda_0 + \epsilon]$. Let p_k denote the number of zeros in (a_k, b_k) of the eigenfunction corresponding to $\lambda = \lambda_k$. Since $\delta(\lambda_0) = 0$, the sequence $p_1, p_2, \cdots$ is bounded. Hence it can be supposed that $p_k = p$ is independent of k; otherwise the sequence $(a_1, b_1), (a_2, b_2), \cdots$ is replaced by a subsequence. If $\mu_p(a, b)$ denotes that point of $S_a{}^b$, which corresponds to an eigenfunction having exactly p zeros on (a, b), Sturm's comparison theorem shows that $\mu_p(a', b') < \mu_p(a, b)$ whenever $a' < a$ and $b' > b$. Since $\lambda_k = \mu_p(a_k, b_k)$, it follows that $\lambda_1 > \lambda_2 > \cdots > \lambda_0 - \epsilon$. The assertion to be proved is an obvious consequence.

(ii) The proof used in (ii) of **3** shows that if $n(\lambda_0 - \epsilon, \lambda_0 + \epsilon) \geqq 1$, then $m(\lambda_0 - \epsilon, \lambda_0 + \epsilon) \geqq 2$; cf. [2], p. 916. The converse statement follows from Sturm's comparison theorem.

(iii) The equivalence of (4) and (4′) in this case can be shown by proving that $m(\lambda_0 - \epsilon, \lambda_0 + \epsilon) \geqq 3$ is necessary and sufficient for λ_0 to be in S. The sufficiency follows from an obvious modification of the construction used in part (iii) of **3**. To show the necessity, assume that λ_0 is in S. Then it belongs to both S^c and S_c, where $t = c$ is a zero of an eigenfunction corresponding to λ_0. Thus, for any $\epsilon > 0$, there exist arbitrarily large numbers $-a, b$ such that $S_a{}^c$ and $S_c{}^b$ each have at least two points in $[\lambda_0 - \epsilon, \lambda_0 + \epsilon]$; cf. [2], p. 918. Then $S_a{}^b$ has at least three points in $[\lambda_0 - \epsilon, \lambda_0 + \epsilon]$ by Sturm's comparison theorem. This completes the proof.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] P. Hartman, "Differential equations with non-oscillatory eigenfunctions," *Duke Mathematical Journal*, vol. 15 (1948), pp. 697-709.

[2] ———, "A characterization of the spectra of one dimensional wave equations," *American Journal of Mathematics*, vol. 71 (1949), pp. 915-920.

[3] ——— and C. R. Putnam, "The least cluster point of the spectrum of boundary value problems," *ibid.*, vol. 70 (1948), pp. 849-855.

[4] ——— and A. Wintner, "A separation theorem for continuous spectra," *ibid.*, vol. 71 (1949), pp. 650-662.

[5] E. Hellinger, *Die Orthogonalinvarianten quadratischer Formen von unendlichen Variablen* (Inaugural-Dissertation), Göttingen 1907.

[6] K. Kodaira, "The eigenvalue problem for ordinary differential equations of the second order and Heisenberg's theory of S-matrices," *American Journal of Mathematics*, vol. 71 (1949), pp. 921-945.

[7] E. C. Titchmarsh, *Eigenfunction expansions associated with second order differential equations*, Oxford (1946).

[8] H. Weyl, "Ueber gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürlicher Funktionen," *Mathematische Annalen*, vol. 68 (1910), pp. 220-269.

[9] ———, "Ueber gewöhnliche Differentialgleichungen mit singulären Stellen und ihre Eigenfunktionen," *Göttinger Nachrichten* (1910), pp. 442-467.

[10] A. Wintner, "Stability and spectrum in the wave mechanics of lattices," *Physical Review*, vol. 72 (1947), pp. 81-82.

ON CUBE-COVERINGS OF n-SPACE.*

By L. G. Peck.

It was shown by Hajos[1] in 1941 that in every covering of Euclidean n-space by unit cubes whose edges are parallel to the coordinate axes and whose centers form a lattice (modul) there is a pair of cubes with a complete $(n-1)$-dimensional face in common. The present note contains the results of an attempt to generalize the above theorem by omitting the condition that the centers of the cubes form a lattice.

$\mathfrak{R}_n$ will denote the n-dimensional vector space over the real field; small German letters will be used for elements of $\mathfrak{R}_n$, and the corresponding Latin letters with subscripts for their components (*e. g.*, $\mathfrak{x} = (x_1, \cdots, x_n)$, $\mathfrak{a}_1 = (a_{11}, \cdots, a_{1n})$). The unit vectors, $\mathfrak{e}_1, \cdots, \mathfrak{e}_n$, are defined by the equation $e_{ij} = 0$ or 1, according as $i \neq j$ or $i = j$. There will also be occasion to use the set Δ_n of n-tuples $\delta = (\delta_1, \cdots, \delta_n)$, where $\delta_i = \pm 1 (i = 1, \cdots, n)$. Note that Δ_n contains 2^n elements.

The following definition is introduced to make precise the notion of a cube-covering of n-space: *A subset* $\mathfrak{A}$ *of* $\mathfrak{R}_n$ *will be called an n-grill if, for every* $\mathfrak{x} \varepsilon \mathfrak{R}_n$, *there exists a unique* $\mathfrak{a} \varepsilon \mathfrak{A}$ *such that* $0 \leqq x_i - a_i < 1$, $(i = 1, \cdots, n)$.

It is clear how one can associate a cube-covering with every n-grill, and an n-grill with every cube-covering. Notice that the defining property of a grill is unaltered by translation, so that there is no loss of generality in assuming that a particular point of the grill is the origin $\mathfrak{o} = (0, \cdots, 0)$.

A property (P) of grills, which will be used repeatedly in the sequel, is as follows:

If $\mathfrak{A}$ *is an n-grill, if* $\mathfrak{a}, \mathfrak{a}' \varepsilon \mathfrak{A}$, *and if* $| a_i - a_i' | < 1$, $(i = 1, \cdots, n)$, *then* $\mathfrak{a} = \mathfrak{a}'$.

Proof. Let $x_i = \max(a_i, a_i')$ for $i = 1, \cdots, n$, and apply the definition of $\mathfrak{A}$.

---

* Received February 9, 1950.

[1] Georg Hajos, "Über einfache und mehrfache Bedeckung des n-dimensionalen Raumes mit einem Würfelgitter," *Mathematische Zeitschrift*, vol. 47 (1941). The reader should also consult papers by Keller, *Journal für die Reine und Angewandte Mathematik*, vol. 163 (1930), and Perron, *Mathematische Zeitschrift*, vol. 46 (1940).

It is now possible to prove a lemma, which, interpreted geometrically, states that the fundamental half-open cube which is used in the covering may be constructed from a closed cube by omitting from it any set of 2^{n-1} non-parallel $(n-1)$-dimensional faces. (The difficulty of making this last statement clearer without departing from geometrical terminology accounts for the introduction of the concept of a grill.)

LEMMA. *Let* $\mathfrak{A}$ *be an* n*-grill,* $\mathfrak{x} \varepsilon \mathfrak{R}_n$*, and* $\delta \varepsilon \Delta_n$*. Then there exists a unique* $\mathfrak{a} \varepsilon \mathfrak{A}$ *such that* $0 \leqq x_i + \delta_i a_i < 1$; $(i = 1, \cdots, n)$.

Proof. The uniqueness of $\mathfrak{a}$ follows from (P). For the existence, we can find (by definition) elements, $\mathfrak{a}_1, \mathfrak{a}_2, \mathfrak{a}_3, \cdots$, of $\mathfrak{A}$ which satisfy

$$\left.\begin{array}{ll} 0 \leqq x_i - a_{ki} < 1 & \text{if } \delta_i = -1 \\ 0 \leqq (-x_i - 1 - 1/k) - a_{ki} < 1 & \text{if } \delta_i = +1 \end{array}\right\} \quad (k = 1, 2, \cdots).$$

Since $| a_{ki} - a_{li} | < 2$ $(i = 1, \cdots, n;\ k, l = 1, 2, \cdots)$, it follows from (P) that there is only a finite number of distinct elements in the sequence $\mathfrak{a}_1, \mathfrak{a}_2, \cdots$. Also, if $\mathfrak{a}_k = \mathfrak{a}_l = \mathfrak{b}$, and $k < l$, we have $0 \leqq x_i - b_i < 1$ if $\delta_i = -1$, and $1/k \leqq (-x_i - 1) - b_i < 1 + (1/l)$ if $\delta_i = +1$. Therefore, if $k < m < l$, we have $| b_i - a_{mi} | < 1$, $(i = 1, \cdots, n)$, and $\mathfrak{a}_m = \mathfrak{b}$ by virtue of (P). Thus, for a suitably large K, $\mathfrak{a}_K = \mathfrak{a}_{K+1} = \cdots = \mathfrak{a}$, say. This $\mathfrak{a}$ clearly satisfies the conditions of the lemma.

With the above preparation, it will now be shown that in any cube covering of the type under consideration each cube has a vertex in common with some other cube. This statement, it should be remarked, is no longer true if the word "vertex" is replaced by the word "edge," trivial counter-examples being available in three dimensions.[2]

THEOREM. *Let* $\mathfrak{A}$ *be an* n*-grill and* $\mathfrak{p} \varepsilon \mathfrak{A}$*. Then there exist numbers,* $\epsilon_1, \cdots, \epsilon_n$*, not all zero, such that* $\epsilon_i = 0$ *or* 1 $(i = 1, \cdots, n)$ *and* $\mathfrak{p} + \epsilon_1 \mathfrak{e}_1 + \cdots + \epsilon_n \mathfrak{e}_n \varepsilon \mathfrak{A}$.

Proof. Without loss of generality, take $\mathfrak{p} = \mathfrak{o}$. Denote by $\mathfrak{A}_0$ the set of $\mathfrak{a} \varepsilon \mathfrak{A}$ which satisfy $0 \leqq a_i \leqq 1$ $(i = 1, \cdots, n)$. If $\delta \varepsilon \Delta_n$, the preceding lemma assures the unique existence of $\mathfrak{a}(\delta) \varepsilon \mathfrak{A}$ satisfying $0 \leqq (1 - \delta_i)/2 + \delta_i a_i(\delta) < 1$; $(i = 1, \cdots, n)$, *i. e.*, $0 \leqq a_i(\delta) < 1$ when $\delta_i = +1$, and

---

[2] Consider the 3-grill whose elements are triples of integers (x, y, z) with the following exceptions: when $x = 0$, $y = \pm 1$, take the third component of the form $z + \frac{1}{2}$, z integral; when $y = 0$, $z = \pm 1$, take the first component of the form $x + \frac{1}{2}$, x integral; when $z = 0$, $x = \pm 1$, take the second component of the form $y + \frac{1}{2}$, y integral. $(0, 0, 0)$ is an element of this grill, but the associated cube has no edge in common with any other cube of the covering.

$0 < a_i(\delta) \leqq 1$ when $\delta_i = -1$. The function $\mathfrak{a}(\delta)$ thus maps each element of Δ_n onto an elements of $\mathfrak{A}_0$. Consider the inverse mapping. If $\mathfrak{b} \varepsilon \mathfrak{A}_0$, we will have $\mathfrak{b} = \mathfrak{a}(\delta)$ on condition that $0 \leqq b_i < 1$ when $\delta_i = +1$, and $0 < b_i \leqq 1$ when $\delta_i = -1$.

In other words, $\delta_i = +1$ when $b_i = 0$, $\delta_i = -1$ when $b_i = 1$, and finally $\delta_i = \pm 1$ when $0 < b_i < 1$. Consequently, *if* $\zeta(\mathfrak{b})$ *denotes the number of components,* b_i, *of* $\mathfrak{b}$ *which satisfy* $0 < b_i < 1$, then the inverse of the mapping $\mathfrak{a}(\delta)$ maps each element $\mathfrak{b} \varepsilon \mathfrak{A}_0$ onto $2^{\zeta(\mathfrak{b})}$ elements of Δ_n. Thus, $\Sigma 2^{\zeta(\mathfrak{b})} = 2^n$, where the summation runs over every $\mathfrak{b} \varepsilon \mathfrak{A}_0$. Since $\zeta(\mathfrak{o}) = 0$, this means that $\Sigma' 2^{\zeta(\mathfrak{b})} = 2^n - 1$ where the $'$ refers to the omission of $\mathfrak{b} = \mathfrak{o}$. Consequently $\Sigma' 2^{\zeta(\mathfrak{b})}$ is odd and so $\zeta(\mathfrak{b}) = 0$ for some $\mathfrak{b} \varepsilon \mathfrak{A}_0$, $\mathfrak{b} \neq \mathfrak{o}$. This completes the proof.

THE JOHNS HOPKINS UNIVERSITY.

ON THE LOCAL UNIQUENESS OF GEODESICS.*

By PHILIP HARTMAN.

1. Statement of the theorems. Let $E(u, v)$, $F(u, v)$, $G(u, v)$ be defined in a neighborhood of $(u, v) = (0, 0)$ in such a way that

$$ds^2 = E du^2 + 2F du dv + G dv^2 \tag{1}$$

is a positive-definite element of arc-length. If E, F, G are of class C^1, then the Christoffel symbols Γ^i_{jk} exist and are continuous functions of (u, v) and so the differential equations

$$u''_i + \Gamma^i_{jk} u'_j u'_k = 0 \text{ for } i = 1, 2, \qquad (u_1, u_2) = (u, v) \tag{2}$$

are meaningful and define the geodesics associated with (1). Professor Wintner has raised the question of whether or not geodesics are uniquely determined (locally) by (2) and by a set of initial conditions (even though the coefficient functions Γ^i_{jk} are not subject to a "uniqueness" restriction, say of the Lipschitz type). It turns out that the answer to this question is of a very surprising nature.

On the one hand, if it is only assumed, as above, that E, F, G are of class C^1, then the solutions of (2) are not in general uniquely determined by initial conditions. On the other hand, if (1) is the first fundamental form of a (2-dimensional) surface of class C^2 in a 3-dimensional Euclidean space, then the solutions of (2) are unique. This last statement can be reworded in the following manner:

THEOREM 1. *On a surface of class C^2, every geodesic is uniquely determined (locally) by its initial conditions.*

The proof of Theorem 1 will depend on the circumstance, pointed out by Weyl [5], pp. 42-44 (cf. also van Kampen [4], p. 135), that the theorem of Gauss-Bonnet and/or the Theorema Egregium (in a suitable form) holds on surfaces of class C^2.

Let X denote the vector (x, y, z) and let $X = X(u, v)$ be of class C^n, where $n \geqq 1$, on a (u, v) domain and such that the vector product $(X_u, X_v) \neq 0$. Then $X = X(u, v)$ is called a surface of class C^n.

The proof of Theorem 1 can be refined to show that it is possible to introduce geodesic polar coordinates (r, ϕ) at any point on a surface of class

---

* Received December 28, 1949.

C^2 and that the resulting transformation $(r, \phi) \to (u, v)$ is of class C^1. In the standard textbooks on differential geometry, it is assumed that the surface $X = X(u, v)$ is of class C^3 in order to introduce geodesic polar coordinates. Their result, under that assumption, is merely that the transformation $(r, \phi) \to (u, v)$ is of class C^1. The proof below will show that if the surface is of class C^n, where $n \geqq 2$, then the mapping $(r, \phi) \to (u, v)$ is of class C^{n-1}.

THEOREM 2. *To every point P on a surface $X = X(u, v)$ of class C^n, where $n \geqq 2$, there corresponds a neighborhood N of P with the property that, for any point $Q \neq P$ in N, there exists one and only one geodesic (in N) joining P and Q. Let $u = u(r, \phi)$, $v = v(r, \phi)$ be the point reached at a distance r from P along a geodesic determined by the angle ϕ from some fixed direction. Then $u = u(r, \phi)$, $v = v(r, \phi)$ are of class C^{n-1} on some circle $0 \leqq r < r_0$, $0 \leqq \phi < 2\pi$ (and have a non-vanishing Jacobian if $r \neq 0$.).*

THEOREM 2*. *The element of arc-length in the (r, ϕ)-coordinates is*

$$ds^2 = dr^2 + g d\phi^2, \tag{3}$$

where $g = g(r, \phi)$ is of class C^{n-2} (and such that $g(0, \phi) \equiv 0$, while $g_r(r, \phi)$ exists and satisfies $g_r(0, \phi) \equiv 0$ even if $n = 2$).

Similar theorems hold for geodesic parallel coordinates.

2. An example of non-uniqueness. It will first be shown that the solutions of (2) need not be unique. To this end, let $h(v)$ be a positive function of class C^1 (for sufficiently small $|v|$). Choose $E(u, v) = G(u, v) = h(v)$ and $F(u, v) \equiv 0$, so that (1) becomes $ds^2 = h(du^2 + dv^2)$. If u is the independent variable, then the system of equations (2) reduces to the single equation

$$2d^2v/du^2 = (1 + (dv/du)^2)H_v, \text{ where } H = \log h, \tag{4}$$

(the subscript v denotes differentiation with respect to v); cf., e. g., the last equation on p. 403 in Darboux [1]. It will be shown that the function $h(v)$ can be chosen so that both $v \equiv 0$ and $v = u^3$ are solutions of (4).

In order that $v \equiv 0$ be a solution, it is sufficient to choose $h(v)$ so as to satisfy $H_v(0) = 0$ (i. e., $h_v(0) = 0$). On the other hand, if (4) is multiplied by dv/du, a quadrature shows that $H(v) = \log(1 + v_u^2)$ or $h(v) = 1 + v_u^2$. If $v = u^3$, then $v_u = 3u^2 = 3v^{2/3}$. In order that $v = u^3$ be a solution of (4), it is sufficient to choose $h(v) = 1 + 9v^{4/3}$. This function $h(v)$ also satisfies $h_v(0) = 0$, and so both $u = 0$ and $u = v^3$ are solutions of (4).

3. Proof of Theorem 1. Let $X = X(u, v)$ be a surface of class C^2 defined in a neighborhood of $(u, v) = (0, 0)$. It will be shown that the initial data $(u, v) = (0, 0)$, $(u', v') = (1, 0)$ determine a unique geodesic. Suppose, if possible, that there are two distinct geodesics satisfying these initial conditions. On both of them, the variable u can be used as parameter. Let them be given by $v = v_1(u)$ and $v = v_2(u)$, respectively, for small $|u|$. Since u can be replaced, if necessary, by $-u$, it can be supposed that $v_1(u) \neq v_2(u)$ for some small positive u. It can also be supposed that

$$E(0,0) = G(0,0) = 1, \qquad F(0,0) = 0. \tag{5}$$

If $\mu > 0$ is sufficiently small, consider the closed curve $S = S_\mu$ which goes from $(0, 0)$ to $(\mu, v_1(\mu))$ along the first geodesic $(v = v_1(u))$, then to $(\mu, v_2(\mu))$ along the parameter curve $u = \mu$, then back to $(0, 0)$ along the second geodesic $(v = v_2(u))$. Although S need not be a simple closed curve, it is clear that one can define the interior $T = T_\mu$ of S, and that Green's theorem is applicable. Hence, the theorem of Gauss-Bonnet gives

$$\int_S \kappa\, ds + \int\int_T K\, d\sigma + \pi + \alpha + \beta = 2\pi,$$

where $K = K(u, v)$ is the Gaussian curvature of the surface at (u, v); κ denotes the geodesic curvature along S; ds is the element of arc-length along S; $d\sigma = (EG - F^2)^{\frac{1}{2}} du\, dv$ is the element of area on T; finally π, α, β are the oriented exterior angles at the respective vertices $(0, 0)$, $(\mu, v_1(\mu))$, $(\mu, v_2(\mu))$ of S. Cf. the proof of the Gauss-Bonnet theorem given by van Kampen [4].

Since the geodesic curvature along geodesics is 0, the last equation reduces to

$$(\pi - \beta) - \alpha = \int_{v_1(\mu)}^{v_2(u)} \kappa G^{\frac{1}{2}}(\mu, v)\, dv + \int_0^{\mu}\int_{v_1(u)}^{v_2(u)} K(EG - F^2)^{\frac{1}{2}} dv\, du. \tag{6}$$

Put

$$m(\mu) = \int_0^{\mu} |v'_2(u) - v'_1(u)|\, du. \tag{7}$$

Then $m(\mu) \neq 0$ for small μ and $m(\mu) \geqq |v_2(u) - v_1(u)|$ for $0 \leqq u \leqq \mu$. It is clear that the integrands in (6) are continuous (hence bounded) functions of μ, u, v near $(\mu, u, v) = (0, 0, 0)$; consequently

$$(\pi - \beta) - \alpha = O(m(\mu)). \tag{8}$$

Since $v'_1(0) = v'_2(0) = 0$, both $\alpha = \alpha(\mu)$ and $\pi - \beta = \pi - \beta(\mu)$ tend to the angle between the parameter curves at $(0, 0)$. This angle is $\pm \frac{1}{2}\pi$, since $F(0, 0) = 0$. By interchanging, if necessary, the numeration of v_1 and v_2, it can be supposed that this angle is $+\frac{1}{2}\pi$. Hence,

$$\alpha - (\pi - \beta) \sim \cos(\pi - \beta) - \cos\alpha \text{ as } \mu \to 0. \tag{9}$$

Since α is the angle between the vector $X_u(\mu, v_1(\mu)) + X_v(\mu, v_1(\mu))v'_1(\mu)$ and $X_v(\mu, v_1(\mu))$, it is seen that

$$\cos\alpha = (F + Gv'_1)\{E + 2Fv'_1 + Gv'^2_1\}^{-\frac{1}{2}}G^{-\frac{1}{2}}, \tag{10}$$

where the arguments of E, F, G are $(\mu, v_1(\mu))$ and that of v'_1 is μ. It is also seen that $\cos(\pi - \beta)$ is given by a similar expression, in which $v_1(\mu)$, $v'_1(\mu)$ are replaced by $v_2(\mu)$, $v'_2(\mu)$, respectively.

For a moment, consider μ, $v_1(\mu)$, $v'_1(\mu)$ as independent variables in the expression on the right of (10). The partial derivative of this function with respect to v'_1 at $(\mu, v_1, v'_1) = (0, 0, 0)$ is 1, by (4). Furthermore, since E, F, G are of class C^1, the expression (10) possesses a continuous (hence bounded) partial derivative with respect to $v_1 = v_1(\mu)$. Hence, as $\mu \to 0$,

$$\begin{aligned} &\cos(\pi - \beta) - \cos\alpha \\ &\qquad = (1 + o(1))(v'_2(\mu) - v'_1(\mu)) + O(v_2(\mu) - v_1(\mu)). \end{aligned} \tag{11}$$

Then (7), (8), (9) imply that $v'_2(\mu) - v'_1(\mu) = O(m(\mu))$. By (7), this means that $m'(\mu) = O(m(\mu))$. But this contains a contradiction, since $m(\mu) \not\equiv 0$ and $m(0) = 0$. The contradiction establishes Theorem 1.

4. **Proof of Theorem 2.** As above, let the surface $X = X(u, v)$ be of class C^2 in a vicinity of $(u, v) = (0, 0)$ and let P be the point $(0, 0)$. Assume the normalization (5) for E, F, G. According to Theorem 1, there is a unique geodesic $u = u(r, \phi)$, $v = v(r, \phi)$ satisfying the initial conditions $(u(0, \phi), v(0, \phi)) = (0, 0)$ and $(u_r(0, \phi), v_r(0, \phi)) = (\cos\theta, \sin\theta)$, where r denotes arc-length on the geodesic. The functions $u = u(r, \phi)$, $v = v(r, \phi)$ possess continuous derivatives of second order with respect to r, on some circle $0 \leqq r < r_0$ and $0 \leqq \phi < 2\pi$, since they are solutions of (2) and the solutions of (2) are uniquely determined by their initial conditions. It will first be shown that u, v and u_r, v_r are of class C^1 on some circle $0 \leqq r < r_0$, $0 \leqq \phi < 2\pi$.

Let $0 < \delta < \frac{1}{2}\pi$, then there exists a positive number $u_0 = u_0(\delta)$ such that, if $|\phi| \leqq \delta$, the equation $u = u(r, \phi)$ has a solution for $r = r(u, \phi)$ whenever $|u| \leqq u_0$. Let $v = v(u; \phi)$ denote the result of substituting $r = r(u, \phi)$ into $v = v(r, \phi)$; so that, for fixed ϕ, the arc $v = v(u; \phi)$ is the geodesic determined by the initial conditions $v(0; \phi) = 0$ and $v_u(0; \phi) = \tan\phi$. The proof of Theorem 1 will be modified to show that $v_u(u; \phi)$ (hence $v(u; \phi)$) has a continuous partial derivative with respect to ϕ for $|\phi| \leqq \delta$ and $|u| \leqq u_0$ (when $u_0 > 0$, $\delta > 0$ are sufficiently small). The arc-length, $r = r(u, \phi)$, on a geodesic is given by

$$(12) \qquad r = r(u, \phi) = \int_0^u \{E + 2Fv_u(\mu; \phi) + Gv_u^2(\mu; \phi)\}^{\frac{1}{2}} d\mu,$$

where the arguments of E, F, G are $(\mu, v(\mu; \phi))$, and accordingly is of class C^1.

Since $u = u(r, \phi)$ is determined by the equation $r = r(u, \phi)$ and since $v = v(r, \phi)$ is obtained by substituting $u = u(r, \phi)$ into $v(u; \phi)$, it will follow that $u(r, \phi)$, $v(r, \phi)$ and $u_r(r, \phi)$, $v_r(r, \phi)$ are of class C^1 on some domain $0 \leqq r < r_1$ and $|\phi| \leqq \delta$. A similar procedure can then be applied on domains $0 \leqq r < r_1(\phi_0)$ and $|\phi - \phi_0| \leqq \delta(\phi_0)$, where ϕ_0 is arbitrary. Finally, an application of the Heine-Borel theorem shows that $u_r(r, \phi)$, $v_r(r, \phi)$ are of class C^1 on some circle $0 \leqq r < r_0$ and $0 \leqq \phi < 2\pi$.

Let ϕ be fixed in such a way that $|\phi| \leqq \delta$, and let $\Delta\phi$ satisfy $|\phi + \Delta\phi| \leqq \delta$ and $|\mu| \leqq \mu_0$. In order to use the notation of the last section, put $v_1(u) = v(u; \phi)$ and $v_2(u) = v(u; \phi + \Delta\phi)$. Consider the closed curve S which goes from $(0, 0)$ to $(\mu, v_1(\mu))$ along the geodesic $v = v_1(u)$, then to $(\mu, v_2(\mu))$ along the parameter curve $u = \mu$, then back to $(0, 0)$ along the geodesic $v = v_2(u)$. Let T denote the "interior" of this curve. As above, the theorem of Gauss-Bonnet is applicable. The equation corresponding to (6) is

$$(13) \qquad (\pi - \beta) - \alpha = -\Delta\phi + \int_{v_1(\mu)}^{v_2(\mu)} \kappa G^{\frac{1}{2}} dv + \int_0^{\mu} \int_{v_1(u)}^{v_2(u)} K(EG - F^2)^{\frac{1}{2}} dv\, du,$$

since the exterior angle at the vertex $(0, 0)$ is $\pi - \Delta\phi$, instead of π as above.

The quantity κ occurring in the first integral on the right of (13) is the geodesic curvature of the parameter curve $u = \mu$ at the point (μ, v). By a standard formula, $\kappa = \det(X_v G^{-\frac{1}{2}}, (X_v G^{-\frac{1}{2}})_v G^{-\frac{1}{2}}, N)$, where N is the unit normal to the surface. Hence, $\kappa G^{\frac{1}{2}}$ is $p(\mu, v)$, where $p(u, v) = G^{-1} \det(X_v, X_{vv}, N)$ is a continuous function of (u, v). Let $q(u, v) = K(EG - F^2)^{\frac{1}{2}}$. Then (13) implies that $(\pi - \beta) - \alpha$ equals

$$(14) \qquad -\Delta\phi + \Delta v(\mu)\{p(\mu, v_1) + o(1)\} + \int_0^{\mu} \Delta v(u)\{q(u, v_1(u)) + o(1)\} du,$$

where $\Delta v = v_2(u) - v_1(u) = v(u; \phi + \Delta\phi) - v(u; \phi)$, and the $o(1)$-terms tend to 0 uniformly with respect to u (for $|u| < u_0$) as $\Delta\phi \to 0$.

On the other hand, the left side of (13) can be treated as follows:

First, (9) can be replaced by

$$\alpha - (\pi - \beta) = \{\cos(\pi - \beta) - \cos\alpha\}\{\sin\alpha + o(1)\}^{-1}.$$

The $o(1)$-term refers to $\Delta\phi \rightarrow 0$ and is uniform with respect to u for $|u| \leqq u_0$; also $\sin\alpha = \det(X_u + X_v v_{1u}, X_v, N)/|X_u + X_v v_{1u}|\,|X_v|$. Thus $\sin\alpha = (EG - F^2)^{\frac{1}{2}}\{E + 2Fv_{1u} + Gv_{1u}{}^2\}^{-\frac{1}{2}}G^{-\frac{1}{2}}$, where the arguments of E, F, G are $(\mu, v_1(\mu))$.

Next, (11) can be replaced by

$$\cos(\pi - \beta) - \cos\alpha = (a + o(1))\Delta v_u + (b + o(1))\Delta v$$

where $a = a(\mu, v_1(\mu), v_{1u}(\mu))$, $b = b(\mu, v_1(\mu), v_{1u}(\mu))$ are the partial derivatives of the expression on the right of (10) with respect to v'_1 and v_1, respectively, at $(\mu, v_1, v'_1) = (\mu, v_1(\mu), v_{1u}(\mu))$. Consequently, $(\pi - \beta) - \alpha$ equals

$$(15) \qquad -\{\sin\alpha(\mu) + o(1)\}^{-1}\{(a + o(1))\Delta v_u + (b + o(1))\Delta v\}.$$

It is clear that neither $\sin\alpha$ nor the function a vanishes if the numbers u_0, δ are sufficiently small.

On interchanging the roles of μ and u, the equality of (14) and (15) can be expressed as

$$\Delta v_u/\Delta\phi = A + B\Delta v/\Delta\phi + C\int_0^u D(\mu)\Delta v/\Delta\phi\, d\mu + o(1),$$

an integro-differential equation for $\Delta v/\Delta\phi$, where A, B, C, D are continuous functions of $u, v(u;\phi)$ and $v_u(u;\phi)$ and, for fixed ϕ, are considered as functions of u. (Actually, $A = \sin\alpha/a$; $B = -(p\sin\alpha + b)/a$; $C = -A$; $D = q$.) It follows by a slight modification of the arguments occurring in the theory of differential equation (cf., e. g., [3], pp. 145-146 or 155-158), that the partial derivative $v_\phi(u, \phi)$ exists, is a continuous function and satisfies the integro-differential equation

$$(16) \qquad y' = A + By + C\int_0^u D(\mu)y(\mu)\,d\mu, \qquad ('= d/du),$$

provided that (16) has a unique solution satisfying $y(0) = 0$. On the other hand, it is easy to show (for example, by the method of successive approximations) that if A, B, C, D are continuous functions of u for $|u| \leqq u_0$, then the linear integro-differential equation (16) possesses, for $|u| \leqq u_0$, a unique solution $y = y(u)$ satisfying an arbitrary initial condition.

Hence $v_\phi(u;\phi)$ exists, is continuous and satisfies (16). But (16) implies that $v_{\phi u}(u, \phi)$ exists and is continuous. Consequently, $v_u(u;\phi)$ is of class C^1 for $|u| \leqq u_0$, $|\phi| \leqq \delta$ (if u_0, δ are sufficiently small). This completes the proof of the statement that $u = u(r, \phi)$, $v = v(r, \phi)$ and their partial derivatives of first order with respect to r are of class C^1 on some circle $0 \leqq r < r_0$, $0 \leqq \phi < 2\pi$.

The proof also shows that if the surface $X(u, v)$ is of class C^n, where $n \geqq 2$, then $u(r, \phi)$, $v(r, \phi)$ (and their partial derivatives with respect to r) are of class C^{n-1}.

Theorem 1 and the standard arguments (cf. [1], p. 408) show that $u = u(r, \phi)$, $v = v(r, \phi)$ are functions of the normal coordinates $\xi = r \cos \phi$, $\eta = r \sin \phi$, say $u = U(\xi, \eta)$, $v = V(\xi, \eta)$. It is easily verified that U, V are of class C^1 on the circle $\xi^2 + \eta^2 \leqq r_0^2$. For example, if $\xi^2 + \eta^2 \neq 0$, then $U_\xi(\xi, \eta) = (\xi/r)u_r - (\eta/r^2)u_\phi$; while $u_r = \cos \phi + O(r)$ and $u_\phi = -r \sin \phi + o(r)$, uniformly in ϕ, as $r \to 0$. Hence, U_ξ exists, is continuous at $(\xi, \eta) = (0, 0)$ and has the value 1 there. Similarly, $U_\eta(0, 0) = 0$; $V_\xi(0, 0) = 0$; $V_\eta(0, 0) = 1$. Consequently, the Jacobian $\partial(U, V)/\partial(\xi, \eta)$ is 1 at $(0, 0)$. This proves Theorem 2.

5. Proof of Theorem 2*. Let $X = Y(r, \phi)$ denote the result of substituting $u = u(r, \phi)$, $v = v(r, \phi)$ into $X = X(u, v)$. Clearly, the scalar product $Y_r \cdot Y_r$ is identically 1, since r is arc-length along the geodesic, $\phi = \text{const}$.

It will be shown that $Y_r \cdot Y_\phi \equiv 0$. Under standard restrictions, this is merely the theorem of Gauss concerning orthogonal trajectories to the geodesics issuing from a point. In order to prove it under the assumptions at hand, the standard arguments (cf. [1], p. 410) will be suitably modified.

First, Y_{rr}, $Y_{r\phi}$ exist and are continuous, and so $(Y_r \cdot Y_\phi)_r = Y_{rr} \cdot Y_\phi + Y_r \cdot Y_{r\phi}$ exists. But $Y_{rr} \cdot Y_\phi \equiv 0$, since Y_{rr}, the principal normal to the geodesic $\phi = \text{const.}$, is parallel to the surface normal N and therefore orthogonal to Y_ϕ. Also, $Y_r \cdot Y_r \equiv 1$ implies $Y_r \cdot Y_{r\phi} \equiv 0$. Hence, $(Y_r \cdot Y_\phi)_r \equiv 0$, so that $Y_r \cdot Y_\phi$ is independent of r. On the other hand, the identity $Y_\phi(0, \phi) \equiv 0$ implies $Y_r \cdot Y_\phi \equiv 0$.

Hence, in terms of dr and $d\phi$, the element of arc-length has the form (3), where $g = Y_\phi \cdot Y_\phi$. Clearly, $g(0, \phi) \equiv 0$. Also, $g_r(r, \phi)$ exists and equals $2Y_\phi \cdot Y_{r\phi}$; so that $g_r(0, \phi) \equiv 0$. Since $g(r, \phi)$ has the smoothness asserted, the Theorem 2* is proved.

6. Remark. Using the notation of Section **5**, let the surface $X = X(u, v)$ be of class C^2 in a vicinity of $(u, v) = (0, 0)$, and let $X = Y(r, \phi)$ denote the parametrization of the surface in geodesic polar coordinates (at $(u, v) = (0, 0)$). As mentioned above (Section **1**), the Theorema Egregium is valid on surfaces of class C^2. It follows that if S is any (smooth) Jordan curve in the circle $0 \leqq r < r_0$ such that its interior T does not contain the point $r = 0$, then

$$\tfrac{1}{2} \int_S 0\, dr + (g_r/g^{\frac{1}{2}})\, d\phi = -\int\int_T K g^{\frac{1}{2}} dr\, d\phi.$$

(This equation is merely an "integrated" form of the Frobenius formula for the Gaussian curvature when (3) holds; this formula is valid on surfaces of class C^2, as can be seen by approximating $X = X(u, v)$ suitably by smooth functions; cf. [2], Section 7.)

Consider curves S which are the boundaries of regions of the type $(0 <) r_1 \leqq r \leqq r_2$, $\phi_1 \leqq \phi \leqq \phi_2$. If the last formula line is divided by $(r_2 - r_1)(\phi_2 - \phi_1)$, and if $\phi_1, \phi_2 \to \phi$, finally, $r_1, r_2 \to r$ $(\neq 0)$, then it is seen that $g_{rr}(r, \phi)$ exists and is continuous if $r \neq 0$, and that $\partial^2 g^{\frac{1}{2}}/\partial r^2 + K g^{\frac{1}{2}} = 0$ if $r \neq 0$.

Although $g(r, \phi)$ is smoother, as a function of r, than is to be expected, it remains an open question as to whether or not $g(r, \phi)$ possesses a partial derivative with respect to ϕ (which is to be expected and hoped for; otherwise, in terms of r, ϕ, it is impossible to consider the usual equations for geodesics). It is easily verified that the orthogonal trajectory $r =$ const. has a parametric representation of class C^2 in terms of its arc-length; however, it is not clear that its parametric representation in terms of ϕ is just as smooth.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] G. Darboux, *Leçons sur la théorie générale des surfaces*, vol. II, Paris (1894).

[2] P. Hartman and A. Wintner, "On the classical embedding problem in differential geometry," *American Journal of Mathematics*, vol. 72 (1950), pp. 553-564.

[3] E. Kamke, *Differentialgleichungen reeller Funktionen*, Leipzig (1930).

[4] E. R. van Kampen, "The theorems of Gauss-Bonnet and Stokes," *American Journal of Mathematics*, vol. 60 (1938), pp. 129-138.

[5] H. Weyl, "Ueber die Bestimmung einer geschlossenen konvexen Fläche durch ihr Linienelement, " *Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich*, vol. 61 (1916), pp. 40-72.

ON THE HÖLDER RESTRICTIONS IN THE THEORY OF PARTIAL DIFFERENTIAL EQUATIONS.*

By AUREL WINTNER.

1. In the following considerations, which deal with solutions u, on an (x, y)-set D, of the partial differential equation

$$\Delta u + f(x, y) u = 0, \tag{1}$$

$(\Delta u = u_{xx} + u_{yy})$, everything will depend on a precise definition of "a solution u on D," as follows:

(I) The (x, y)-set D will be required to be open (and connected or, for that matter, simply-connected; either of the latter restrictions will be immaterial, since there will no loss in generality in assuming that D is "small"). The open set D will be referred to as a domain if it is not the empty set.

(II) The coefficient function, f, of (1) will be assumed to be (real-valued and) continuous on a given D.

(III) A function u, defined on D, will be called a solution, on D, of a differential equation if and only if all those derivatives of u which occur in differential equation exist, and satisfy the differential equation, at every point of D.

Remark. In the sequel it will be essential that the notion of "a u which satisfies the differential equation almost everywhere on D," a notion first used by Lichtenstein and further developed by Schauder, Leray, and others, is entirely ruled out by (I) and (III). In fact, the ejection of an unspecified zero set (or, for that matter, of an enumerable set) replaces the domain D by a set which, instead of being again a domain, can fail to contain any domain as a subset. Correspondingly, the question is *not* an issue just in a Hilbert space (such as the u-space $\operatorname{grad} u \subset L^2$ on D).

2. If (III) is applied to (1), all that follows for the smoothness of u is the existence of the derivatives u_{xx}, u_{yy} (but not that of u_{xy} and u_{yx}, and still less their equality) at every point of D. This implies the existence of u_x and u_y, but it fails to imply the existence of a complete differential du;

---

* Received January 24, 1950.

as a matter of fact, not even the continuity of $u(x,y)$ can be concluded. In fact, solutions of (1) which are discontinuous on a dense set are compatible with (III) even in the case, $f \equiv 0$, of Laplace's equation. This can be seen as follows:

Define on the (x,y)-plane a function $u(x,y)$ by placing $u(0,0)=0$, and $u(x,y) = -2xy/(x^2+y^2)^2$ if $(x,y) \neq (0,0)$. Then, since $u(x,0)$ and $u(0,y)$ vanish identically, u_{xx} and u_{yy} exist and vanish, hence they satisfy $\Delta u = 0$, at the point $(x,y)=(0,0)$. Except for this point, $u(x,y)$ is a regular harmonic function, since it is the imaginary part of $1/(x+iy)^2$. Consequently, $u(x,y)$ is a solution of $\Delta u = 0$ on D in the sense of (III), where D is the entire (x,y)-plane, although $u(x,y)$ is discontinuous at $(0,0)$, since it is unbounded near $(0,0)$. In order to "condense" this isolated "singularity," it is sufficient to put $u_n(x,y) = u(x-x_n, y-y_n)$ and $u^*(x,y) = \Sigma\, a_n u_n(x,y)$, where (x_1,y_1), (x_2,y_2), $\cdots$ is a sequence of points which is dense on the (x,y)-plane, and $a_1, a_2, \cdots$ is a sequence of constants, where it is understood that the elements of these two sequences must be chosen by an appropriate induction. What then results is a function u^* which, in the sense of (III), is a solution of Laplace's equation on D (with D the entire (x,y)-plane), but which is such as to be discontinuous at every point of a set of second category (in fact, $u^*(x,y)$ can be unbounded on every circle $(x-x_n)^2+(y-y_n)^2<\epsilon$, where $n=1,2,\cdots$).

In order to avoid pathological solutions of this type, only those solutions of (1) will be considered which, besides being solutions in the sense of (I)-(III), are continuous functions. This additional restriction of the smoothness required by (III) is the more natural as every continuous solution u of $\Delta u = 0$ on D is of class C'' (hence, a regular harmonic function) on D; cf. [8] or [5].

3. It is known (for references, cf. [3], pp. 1294-1295) that if D is small enough and if J is a sufficiently smooth Jordan curve contained in D, then, with reference to every smooth continuous function assigned on J, the first boundary-value problem for the interior of J (with a $u(x,y)$ which is continuous within and on J) is solvable for (1), provided that the coefficient function satisfies a Hölder condition,

$$|f(x',y')-f(x'',y'')| \leqq \text{const.}\{(x'-x'')^2+(y'-y'')^2\}^{\frac{1}{2}\theta}, \qquad (0<\theta<1).$$

It will be shown that the situation becomes entirely different if this proviso is omitted (in this regard, cf. [4]) and just the continuity of the coefficient functions is required. In fact, there will be constructed a (real-valued) $f(x, y)$ having the following property:

(*) Although $f(x, y)$ is continuous on the entire (x, y)-plane, there does not exist *any* domain D (or, equivalently, *any* circle D: $(x-x_0)^2+(y-y_0)^2<r^2$, with some (x_0, y_0) and some $r>0$) on which D the differential equation (1) possesses *some* continuous solution, $u(x, y)$, distinct from the trivial solutions, $u(x, y) \equiv 0$.

Accordingly, *every* boundary value problem belonging to (1) becomes unsolvable for *every* Jordan curve J, except when the boundary condition on J is such as to be solved by $u(x, y) \equiv 0$ (within J, hence on J). In addition, it is possible to choose a continuous $f(x, y)$ as follows:

(**) The boundary-value problem

$$u+\{\lambda+f(x, y)\}u=0, \qquad u=0 \text{ on } J, \tag{2}$$

becomes meaningless for every circle J: $(x^2-x_0)^2+(y-y_0)^2=r^2$, since it is possible that (2) has, for *no* value of the parameter λ, and on *no* (x, y)-domain, *any* continuous solution distinct from $u(x, y) \equiv 0$, even though $f(x, y)$ is continuous throughout.

This refinement, (**), of (*) will follow from the proof (*).

4. Quite another consideration, one based simply on the equation,

$$y'=-f(x, y), \qquad ({}'=d/dx), \tag{a}$$

of the characteristic curves, $y=y(x)$, of

$$u_x+f(x, y)u_y=0, \tag{b}$$

shows that similar possibilities can occur for linear, homogeneous, partial differential equations of *first* order also. For the inhomogeneous extension, $u_x+f(x, y)u_y=g(x, y)$, of the homogeneous equation, (b), the construction of a counter-example is more straightforward, since, according to Perron [6], p. 550, it can be made to depend on the choice of the continuous function $g(x, y)$, whilst $f(x, y) \equiv 1$, and this would not do in (b).

Only solutions of class C' will be considered for (b). Then what corresponds to the assertions of Section **3** is as follows:

(†) There exist on the (x, y)-plane real-valued, continuous functions $f(x, y)$ having the following property: There does not exist *any* domain D on which D the differential equation (4) possesses *some* solution of class C', say $u=u(x, y)$, distinct from the trivial solutions, $u(x, y)=\text{const.}$

First, if $f(x, y)$ is unspecified, and if $y=y(x)$ is any solution curve of (a), it is clear that $du(x, y(x))/dx$ vanishes, identically in x, whenever $u(x, y)$ has a complete differential (e. g., bounded partial derivatives u_x, u_y) and is a solution of (b) (on an (x, y)-domain, D, which contains the curve $y = y(x)$). Hence, if a point (x_0, y_0) of D issues more than one characteristic, say K_1: $y = y_1(x)$ and K_2: $y = y_2(x)$, where, without loss of generality, $y_1(x) < y_2(x)$ holds for small positive $x - x_0$, then $u(x, y)$ has the same value, $u(x_0, y_0)$, at every point of $K_1 + K_2$. But it is well-known (Osgood) that if (x^*, y^*) is any point, close enough to (x_0, y_0), which is contained in the "horn" bordered by K_1 and K_2, then (a) possesses at least one solution path, K: $y = y(x)$, which joins (x^*, y^*) with (x_0, y_0). Consequently, $u(x, y)$ is constant on the entire domain of the "horn."

Choose the real-valued, continuous $f(x, y)$ in such a way that every point of the (x, y)-plane becomes a point (x_0, y_0), that is, a point of non-uniqueness, of (a) (such functions f exist; cf. [1]). Then every point (x, y) will be a boundary point of some domain $D = D_{xy}$ (viz., the open set supplied by a "horn") on which domain the function $u(x, y)$ is constant; so that $u_x \equiv 0 \equiv u_y$ on $D = D_{xy}$. Hence, if $u(x, y)$ of class C', then $u_x \equiv 0 \equiv u_y$ holds throughout. This proves (†).

4 bis. Clearly, the coefficient function, $f(x, y)$, of the example, (b), proving (†) cannot vanish identically on any (x, y)-domain. Hence, $f(x, y)$ will not vanish at some point and therefore, by continuity, on some (x, y)-square. It can therefore be assumed that $f(x, y)$ is positive (or, equivalently, that $f(x, y)$ is negative) throughout the square $(|x| \leqq 1, |y| \leqq 1)$. This leads to the following corollary of (†):

(† bis) There exist on the square $(|x| \leqq 1, |y| \leqq 1)$ continuous functions $f(x, y) > 0$, $g(x, y) < 0$ having the property that the binary linear homogeneous system

$$u_x = f(x, y) u_y + 0 \cdot v_y, \qquad v_x = 0 \cdot u_y + g(x, y) v_y$$

(which, in view of $f(x, y) \neq g(x, y)$, is a *hyperbolic* system) has on no open set contained in $(|x| \leqq 1, |y| \leqq 1)$ any solution of class C' distinct from the trivial solution: $u(x, y) =$ const., $v(x, y) =$ Const.

In fact, the system in the last formula line splits into an equation for u alone and one for v alone in such a way that (†) is applicable to each of them.

5. The proof of (*) will depend on certain facts concerning the solutions of Poisson's equation,

$$\Delta u = g(x, y), \tag{3}$$

and its formal solutions,

$$(4) \qquad u_D(x,y) = (2\pi)^{-1} \int\int_D g(\xi,\eta) \log r \, d\xi d\eta,$$

where $r^2 = (\xi - x)^2 + (\eta - y)^2$. The first of these facts is as follows:

(i) If $g(x,y)$ is continuous and bounded on a bounded domain D, then (3) cannot have any continuous solution on D unless the function (4) represents a solution of (3) on D.

In order to ascertain (i), let $\Delta^* v = \Delta^* v(x,y)$ denote, for any point (x,y) and for any function v defined in a vicinity of (x,y), the limit of

$$h^{-2}\{v(x+h,y) + v(x,y+h) + v(x-h,y) + v(x,y-h) - 4v(x,y)\}$$

as $h \to 0$, provided that this limit exists. It is known (cf. [8], p. 147) that if a function $v(x,y)$ is continuous on a domain D and is such that $\Delta^* v$ exists and vanishes at every point (x,y) of D, then $v(x,y)$ is a regular harmonic function on D. Furthermore, it is readily seen (cf. [8], p. 149) that if $v(x,y)$ is a function for which Δv exists at a point (x_0,y_0) (in the sense that both $u_{xx}(x_0,y_0)$ and $u_{yy}(x_0,y_0)$ exist, implying that $u_x(x,y_0)$, $u_y(x_0,y)$ exist for small $|x - x_0|$, $|y - y_0|$, respectively), then $\Delta^* v$ exists, and equals Δv, at (x_0,y_0). Finally, Zaremba [8] has shown that if $g(x,y)$ is continuous on a bounded domain D, then the function (4) satisfies, at every point of D, the following variant of (3):

$$(3 \text{ bis}) \qquad \Delta^* u_D(x,y) = g(x,y)$$

(the existence of $\Delta^* u_D$ at every point of D is the essential part of the statement).

Clearly, these three facts imply the truth of (i), simply because the logarithmic potential (4) is a continuous function on D. As a matter of fact, (4) always is a function of class C' on D. Hence, the three facts leading to (i) also lead to the following lemma (cf. [8]):

(ii) If $g(x,y)$ is continuous on a domain D, and if (3) has on D a continuous solution $u(x,y)$, then $u(x,y)$ is of class C' on D.

In fact, it is clear from the proof of (i) that the situation is as follows:

(iii) If $g(x,y)$ is continuous on a bounded domain D, then (3) either does not have any continuous solution $u(x,y)$ on D or, if it has such a solution, its most general such solution is given by $u = u_D + v$, where u_D is defined by (4) and v denotes an arbitrary function which is (regular) harmonic on D.

6. The following analogue of (*) will now be proved:

(§) It is possible to choose, on the (x, y)-plane, a continuous function, $g(x, y)$, and a dense set, S, of points in such a way that, if D is any bounded domain, the function (4) fails to have a second derivative $\partial^2 u_D/\partial x^2$ at every point of the common part of D and S.

It follows from (i) that (§) actually is an analogue of (*), with (1) replaced by (3).

In the proof of (§), on which that of (*) will depend, the following criterion of Petrini [7], pp. 132-133, will be needed: If (a, b) is a point of D, and if $E(\epsilon; a, b)$ denotes the circle $r \leqq \epsilon$, where $r^2 = (\xi - a)^2 + (\eta - b)^2$, then the existence of the limit

$$(5) \qquad \lim_{\epsilon \to 0} \iint_{D-E(\epsilon; a, b)} g(\xi, \eta) \partial^2 \log r / \partial \xi^2 \, d\xi d\eta$$

is necessary and sufficient in order that the function (4), where g is continuous, be such as to possess a second derivative $\partial^2 u_D/\partial x^2$ at the point (a, b) (a similar criterion holds for the existence of the other second derivatives of u_D).

Since $r^2 = (\xi - a)^2 + (\eta - b)^2$, it is clear that

$$(6) \qquad \iint_{D-E(\epsilon; a, b)} | \partial^2 \log r / \partial \xi^2 | \, d\xi d\eta \to \infty, \text{ as } \epsilon \to 0,$$

holds for every (a, b). But the integral (6) is the "Lebesgue constant" which, in the "singular integral" occurring in (5), belongs to a fixed $\epsilon > 0$. It follows therefore from (6), by Lebesgue's classical construction (cf. [2], pp. 60-63), that there exists a continuous function $f(x, y) = f_{ab}(x, y)$ for which the limit (5) fails to exist. But it is well-known that Lebesgue's construction can be carried out for a sequence of points simultaneously, if the "Lebesgue constants" (belonging to the parameter, which is the above ϵ) are unbounded (as $\epsilon \to 0$) at each of the points belonging to the sequence.[1] Consequently, there exists a continuous function $g(x, y)$ having the property that the limit (5) fails to exist at each of the points $(a, b) = (a_1, b_1)$, $(a_2, b_2), \cdots$, where (a_n, b_n) is any point of D. Hence the assertion of (§) follows by choosing the dense set, occurring in (§), as follows:

$$(7) \qquad S\colon \; (x_n, y_n), \qquad\qquad (n = 1, 2, \cdots),$$

---

[1] In fact, if the case of a single point, say P, is replaced by that of a sequence, $P_1, P_2, P_3, \cdots$, it is sufficient to replace the sequence $P, P, P, \cdots$, which occurs in Lebesgue's construction, by the sequence P_1; P_1, P_2; P_1, P_2, P_3; $P_1, \cdots$.

where (x_1, y_1), (x_2, y_2), $\cdots$ denote those points of the (x, y)-plane for which both x and y are rational.

7. Besides (§), the following criterion of comparison will be needed in the proof of (*):

(§§) Let $g^1(x, y)$, $g^2(x, y)$ be two continuous, bounded functions on a bounded domain D containing the point (a, b), and let $u_D{}^1(x, y)$, $u_D{}^2(x, y)$ denote the corresponding functions (4). Suppose that $g^1 = Gg^2$, where $G(x, y)$ is a function satisfying $G(a, b) \neq 0$ and having a complete differential $dG(x, y)$ at (a, b) (e.g., continuous derivatives G_x, G_y). Then either both or neither of the functions $u = u_D{}^1$, $u = u_D{}^2$ has a second derivative $\partial^2 u/\partial x^2$ at $(x, y) = (a, b)$.

In fact, if $r^2 = (\xi - a)^2 + (\eta - b)^2 \rightarrow 0$, the existence of dG at (a, b) means that

$$G(\xi, \eta) = G(a, b) + \text{const.}(\xi - a) + \text{Const.}(\eta - b) + o(r),$$

and implies therefore that $G(\xi, \eta) = c + O(r)$, where $c = G(a, b)$. It follows therefore from $g^1 = Gg^2$, where $g^2(\xi, \eta) = O(1)$, that

$$g^1(\xi, \eta) = cg^2(\xi, \eta) + O(r), \text{ where } c \neq 0, \tag{8}$$

since $G(a, b) \neq 0$. In addition, (8) remains true if g^1 and g^2 are interchanged and c is replaced by $1/c$. But the existence of the limit (5) is necessary and sufficient for the existence of the $\partial^2/\partial x^2$-derivative of (4) at $(x, y) = (a, b)$. It follows therefore from (8) that the assertion of (§§) will be proved if it is ascertained that

$$\iint\limits_{D-E(\epsilon;a,b)} O(r)\partial^2 \log r/\partial\xi^2 \, d\xi d\eta \rightarrow 0, \text{ as } \epsilon \rightarrow 0.$$

But this is sure to be true if

$$\iint\limits_{r\leqq 1} r \mid \partial^2 \log r/\partial\xi^2 \mid d\xi d\eta < \infty,$$

and the convergence of the last integral is seen by placing $\xi = a + r\cos\phi$, $\eta = b + r\sin\phi$.

8. In order to prove (*), choose a function g satisfying the requirements of (§) with regard to (3), and put $f = -g$ (or, for that matter, $f = g$). It will be shown that this f satisfies the requirements of (*) with regard to (1). This will also prove that the same f satisfies the requirements of (**) with regard to (2) and to every $\lambda = \text{const.}$; simply because an additive constant has no bearing on the construction leading to the $g(x, y)$ of (§).

For the given $f = -g$, let u be a continuous solution of (1) on a bounded D. The assertion is that $u \equiv 0$ on D.

First, if (1) is written in the form (3), where g represents the product of the two continuous functions $-f$, u, it follows from (i)-(iii) that the given continuous solution, $u(x, y)$, of (1) must be of class C' on D, and that the logarithmic potential (4) belonging to $g = -fu$ must have a second derivative $\partial^2 u_D/\partial x^2$ at every point (x, y) of D. It follows therefore from (§§) that, if $v_D = v_D(x, y)$ denotes the logarithmic potential (4) belonging to $g = -f$, and if (x^*, y^*) is in D, then either $\partial^2 v_D/\partial x^2$ exists at (x^*, y^*) or $u(x^*, y^*) = 0$. But g was chosen in accordance with (§). Hence, $\partial^2 v_D/\partial x^2$ fails to exist at every point of an (x, y)-set, (8), which is dense on D. Consequently, u must be 0 at every point of some dense subset of D. It follows therefore from the continuity of D that $u \equiv 0$ on D.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] M. Lavrientieff, "Sur une équation différentielle du premier ordre," *Mathematische Zeitschrift*, vol. 23 (1925), pp. 97-209.

[2] H. Lebesgue, "Sur les intégrales singulières," *Annales de la Faculté des Sciences de Toulouse*, (3), vol. 1 (1910), pp. 125-127.

[3] L. Lichtenstein, "Neuere Entwicklung der Theorie partieller Differentialgleichungen zweiter Ordnung vom elliptischen Typus," *Encyklopädie der Mathematischen Wissenschaften*, II C 12 (1924), pp. 1277-1334.

[4] ———, "Neue Untersuchungen in der Theorie der linearen partiellen Differentialgleichungen zweiter Ordnung vom elliptischen Typus," *Bulletin de l'Académie Polonaise des Sciences et des Lettres*, ser. A (1931), pp. 571-598.

[5] H. Looman, "Bemerkung zur Definition der harmonischen Funktionen," *Christian Huygens*, vol. 3 (1923-24), pp. 41-42.

[6] O. Perron, "Ueber Existenz und Nichtexistenz von Integralen partieller Differentialgleichungssysteme im reellen Gebiet," *Mathematische Zeitschrift*, vol. 27 (1928), pp. 549-564.

[7] H. Petrini, "Les dérivées premières et secondes du potentiel logarithmique," *Journal de Mathématiques*, ser. 6, vol. 5 (1909), pp. 127-223.

[8] S. Zaremba, "Contribution à la théorie d'une équation fonctionnelle de la physique," *Rendiconti del Circolo Matematico di Palermo*, vol. 19 (1905), pp. 140-150.

ON THE ZETA-FUNCTIONS OF ALGEBRAIC NUMBER FIELDS II.*

By RICHARD BRAUER.

1. It was shown in the first paper[1] that if k is an algebraic number field of a fixed degree n over the field P of rational numbers, then the discriminant d of k, the regulator R of k, and the number h of classes of ideals in k are connected by the asymptotic relation

$$(1) \qquad \log (Rh) \sim \log \sqrt{|d|}$$

for $|d| \to \infty$. In this note, fields k will be considered whose degree n over P is no longer fixed. It will be shown that (1) holds for every sequence of fields k normal over P for which $n/\log |d|$ tends to zero.[2] If k is an arbitrary algebraic number field and if n^* is the degree of the minimal normal field k^* over P which contains k, the set of values

$$\left(\frac{\log (Rh)}{\log \sqrt{|d|}} - 1\right)\frac{1}{n^*}$$

(for all $k \neq \mathrm{P}$) is bounded and has zero as its only limit point. Since $n^* \leqq n!$, this result implies (1) for fields of fixed degree n.

The procedure is similar to that of I. In particular, the method used in **4** is an extension of a method of C. L. Siegel.[3] Except for the proof of the Lemma on p. 244 of I, the present paper is independent of I.

2. We shall denote by $\rho(k)$ the residue at $s = 1$ of the zeta-function $\zeta(s, k)$ of the algebraic number field k.

LEMMA. *There exists an absolute constant c_1 such that the inequality*

$$(2) \qquad \rho(k) \leqq c_1^n (1+\gamma)^n |d|^{1/2\gamma}$$

holds for all algebraic number fields k and all $\gamma \geqq 1$.[4]

---

* Received February 7, 1950; presented to the American Mathematical Society, November 25, 1949.

[1] *American Journal of Mathematics*, vol. 69 (1947), pp. 243-250. We shall quote this paper by I.

[2] As was shown by Minkowski, this quotient $n/\log |d|$ lies below a fixed constant for all algebraic number fields $k \neq \mathrm{P}$.

[3] C. L. Siegel, *Acta Arithmetica*, vol. 1 (1935), pp. 83-86.

[4] We shall use the notation $c_1, c_2, \cdots$ for positive constants which do not depend on the field k. Similarly $c_i(\epsilon)$ will denote a positive constant which depends on a parameter ϵ but is independent of k.

For the proof, compare Landau, *Nachrichten von der Koeniglichen Gesellschaft der Wissenschaften zu Göttingen,* Mathematisch-Physikalische Klasse 1918, pp. 478-488, proof of Lemma 1. If a is the number of real conjugates of k, and $2b$ the number of non-real conjugates, $n = a + 2b$, then the expression

$$| d |^{s/2} 2^{-bs} \pi^{-ns/2} \Gamma(s/2)^a \Gamma(s)^b \zeta(s, k) - \rho(k) \, | d |^{1/2} (2\pi)^{-2b} / s(s-1)$$

is positive for all real s. This follows from Hecke's representation of the expression as an integral.[5] The lemma is obtained by setting $s = 1 + \gamma^{-1}$ and using the inequality

$$\zeta(1 + \gamma^{-1}, k) \leqq \zeta(1 + \gamma^{-1})^n \leqq (1 + \gamma)^n$$

where $\zeta(s) = \zeta(s, \mathrm{P})$ is Riemann's zeta-function.

The value of the residue $\rho(k)$ is given by

$$\rho(k) = 2^n \pi^b Rh / w \sqrt{| d |} \tag{3}$$

where w is twice the number of roots of unity in k if $b > 0$, and $w = 2$ if $b = 0$. We need the following rough estimate of w:

$$w \leqq c_2 n^2. \tag{4}$$

For the proof of (4) we may assume $w > 2$, which implies $b > 0$. Since k contains the roots of unity of order $w/2$, we have $n \geqq \phi(w/2)$ where ϕ denotes the Euler function. Suppose now that $w > 2n^2$. Then $w/2 > n^2 \geqq \phi(w/2)^2$. Since $\phi(q)/q^{1/2} \to \infty$ for $q \to \infty$,[6] it follows that w must lie below a certain constant. This proves (4).[7]

Combination of (2), (3), and (4) shows that

$$\begin{aligned} \log(Rh) - \log\sqrt{| d |} \leqq n \log(c_1(1+\gamma)) &+ (\log\sqrt{| d |})/\gamma \\ &+ 2 \log n + \log c_2. \end{aligned} \tag{5}$$

Since $n/\log | d |$ lies below a fixed constant (cf. footnote 2), it follows for $\gamma = 1$ that there exists a constant c_3 with

$$\log(Rh)/\log\sqrt{| d |} \leqq c_3 \qquad (\text{for } k \neq \mathrm{P}). \tag{6}$$

If k ranges over a sequence of fields for which $n/\log | d |$ tends to zero, then (5) implies that

$$\overline{\lim} \log(Rh)/\log\sqrt{| d |} \leqq 1 + \gamma^{-1}.$$

---

[5] See, for instance, Siegel, *loc. cit.*[3], equation (5).

[6] See, for instance, Hardy and Wright, *An Introduction to the Theory of Numbers,* Oxford 1945, Theorem 327, p. 265.

[7] If we use Landau's inequality $\phi(q) \geqq c\, q/\log\log q$ (c a positive constant, $q > 1$), (5) can be improved to $w \leqq c' n \log\log n + c''$ (c', c'' positive constants, $k \neq \mathrm{P}$).

If $\gamma \to \infty$, it follows that

$$\overline{\lim} \log (Rh)/\log \sqrt{|d|} \leqq 1 \qquad (\text{for } n/\log |d| \to 0). \tag{7}$$

Finally, if k ranges over an infinite sequence of distinct fields, then $|d| \to \infty$ and (5) yields

$$\overline{\lim} \left(\frac{\log (Rh)}{\log \sqrt{|d|}} - 1 \right) \frac{1}{n} \leqq 0 \qquad (\text{for } |d| \to \infty). \tag{8}$$

3. We now prove

THEOREM 1. *Let k be an algebraic field and let K be a normal extension field of k. Denote by d the discriminant of k and by D the discriminant of K. Set $[k:\mathrm{P}] = n$ and $[K:\mathrm{P}] = N$. There exists a constant c_4 such that*

$$\rho(K)/\rho(k) \leqq c_4^{N-n}(1+\gamma)^{N-n}(|D|/|d|)^{1/2\gamma} \tag{9}$$

for all $\gamma \geqq 1$.

Proof. Let $\mathfrak{G}$ be the Galois group of K with regard to k. It was shown in I that if θ is the character of the regular representation of $\mathfrak{G}$ and if χ_0 is the 1-character of $\mathfrak{G}$, we have an equation of the form [8]

$$\theta - \chi_0 = \sum \lambda(j) \psi^*_j \tag{10}$$

where each ψ_j is an irreducible character different from the 1-character of a cyclic subgroup $\mathfrak{A}_j$ of $\mathfrak{G}$, ψ^*_j denotes the character of $\mathfrak{G}$ induced by the character ψ_j of $\mathfrak{A}_j$, and the $\lambda(j)$ are positive rational coefficients.

Let the subgroup $\mathfrak{A}_j$ of the Galois group $\mathfrak{G}$ correspond to the field Z_j; $k \subseteq Z_j \subseteq K$. Then K is cyclic over Z_j and hence it is class field to a certain group $\mathfrak{H}_j$ of ideals in Z_j. On account of the fundamental results of Artin,[9] (10) implies

$$\zeta(s, K)/\zeta(s, k) = \prod_j L(s, \psi^*_j, K/k)^{\lambda(j)}; \tag{11}$$

moreover $L(s, \psi^*_j, K/k)$ is equal to an abelian L-series:

$$L(s, \psi^*_j, K/k) = L(s, \psi_j), \tag{12}$$

where ψ_j on the right is considered as a character of the group of ideal classes $(\bmod\, \mathfrak{H}_j)$ in Z_j. It follows from (11) and (12) for $s \to 1$ that

---

[8] See the Lemma on p. 244 of I. Cf. also Aramata, *Proceedings of the Imperial Academy of Japan*, vol. 9 (1933), pp. 31-34.

[9] *Abhandlungen aus dem Mathematischen Seminar Hamburg*, vol. 3 (1924), pp. 89-108; vol. 8 (1931), pp. 292-306.

$$(13) \qquad \rho(K)/\rho(k) = \prod_j | L(1, \psi_j) |^{\lambda(j)}.$$

The factors on the right are finite since ψ_j is different from the 1-character.

We now need an upper estimate for the factors $| L(1, \psi_j) |$. Let $n(j)$ be the degree of Z_j over P; let $a(j)$ denote the number of real conjugates of Z_j and let $2b(j)$ denote the number of non-real conjugates. If $\mathfrak{f}_j$ is the conductor of ψ_j and if d_j is the discriminant of Z_j, we set $\kappa_j = | d_j |^{1/2} \mathfrak{N}(\mathfrak{f}_j)^{1/2}$ where $\mathfrak{N}(\mathfrak{f}_j)$ denotes the norm taken in Z_j with respect to P. Finally, we set

$$\Gamma(s, \psi_j) = \Gamma((s+1)/2)^q \Gamma(s/2)^{a(j)-q} \Gamma(s)^{b(j)},$$

where q is the number of infinite primes of Z_j appearing in $\mathfrak{f}_j$. Then, for every $s > 1$, the inequality holds[10]

$$\kappa_j 2^{-b(j)} \pi^{-n(j)/2} \Gamma(1, \psi_j) | L(1, \psi_j) | \\ \leqq 2 (\kappa_j 2^{-b(j)} \pi^{-n(j)/2})^s \Gamma(s, \psi_j) \zeta(s, Z_j).$$

Set $s = 1 + \gamma^{-1}$ with $\gamma \geqq 1$. Since

$$\zeta(1 + \gamma^{-1}, Z_j) \leqq \zeta(1 + \gamma^{-1})^{n(j)} \leqq (1 + \gamma)^{n(j)},$$

we obtain an inequality of the form $| L(1, \psi_j) | \leqq c_4^{n(j)} (1 + \gamma)^{n(j)} \kappa_j^{1/\gamma}$ with a positive constant c_4. Now, (13) yields

$$(14) \qquad \rho(K)/\rho(k) \leqq \prod_j c_4^{n(j)\lambda(j)} (1 + \gamma)^{n(j)\lambda(j)} \kappa_j^{\lambda(j)/\gamma}.$$

In order to evaluate the product on the right, we compare the number of zeros of the functions appearing in (11) in the rectangle

$$(15) \qquad 0 \leqq \mathfrak{R}(s) < 1, \qquad 0 < \mathfrak{J}(s) \leqq t$$

(for $t > 0$). If M_j denotes the number of zeros of $L(s, \psi_j)$ in this rectangle, then, by the analogue of the Riemann-von Mangoldt theorem,[11]

$$2\pi M_j = n(j) t \log t + (2 \log \kappa_j - n(j) - n(j) \log 2\pi) t + O(\log t).$$

The same formula holds for the number of zeros of $\zeta(s, K)$ in the rectangle, if we take for $n(j)$ here the degree N of K and for κ_j the value $| D |^{1/2}$. Similarly, in the case of $\zeta(s, k)$, we must take $n(j) = n$ and $\kappa_j = | d |^{1/2}$.

---

[10] This is shown by the method used by Landau, *Mathematische Zeitschrift*, vol. 4 (1919), pp. 152-162, in the proof of Theorem 4. If we set

$$(\kappa_j 2^{-b(j)} \pi^{-n(j)/2})^s \, \Gamma(s, \psi_j) L(s, \psi_j) = \Phi(s),$$

then according to Hecke, $\Phi(s)$ can be written as a sum of two integrals $\Psi_1(s)$ and $\Psi_2(s)$. It follows from the form of these integrals that $| \Psi_i(1) |$ is at most equal to the value obtained from $\Phi(s)$ by replacing $L(s, \psi_j)$ by $\zeta(s, Z_j)$ and taking $s > 1$.

[11] See Landau, *Mathematische Zeitschrift*, vol. 2 (1918), pp. 52-154, Theorem LXXI.

Move the factor $\zeta(s,k)$ in (11) to the right and raise the equation to the l-th power where l is the least common denominator of all $\lambda(j)$. Then, on comparing the number of zeros of both sides in the rectangle (15), we find

$$\begin{aligned} &lNt\log t + l(\log|D| - N - N\log 2\pi)t + O(\log t) \\ &\quad = lnt\log t + l(\log|d| - n - n\log 2\pi)t \\ &\qquad + \sum_j \{t\log t\, n(j) l\lambda(j) + (2\log\kappa_j - n(j) - n(j)\log 2\pi) l\lambda(j) t\}. \end{aligned}$$

This yields [12]

$$N = n + \sum_j n(j)\lambda(j); \tag{16}$$

$$\log|D| = \log|d| + 2\sum_j \lambda(j)\log\kappa_j. \tag{17}$$

Hence

$$\prod_j c_4^{n(j)\lambda(j)}(1+\gamma)^{n(j)\lambda(j)} = c_4^{N-n}(1+\gamma)^{N-n};$$

$$\prod_j \kappa_j^{\lambda(j)/\gamma} = (|D|/|d|)^{1/2\gamma}.$$

If these values are substituted in (14), the inequality (9) is obtained and Theorem 1 is proved.

Remark 1. In the case that K is normal over k, the lemma in **2** may be considered as a special case of Theorem 1.

Remark 2. If $K \neq k$ and we choose $\gamma = \log(|D|/|d|)$ in Theorem 1, we obtain

$$\rho(K)/\rho(k) \leqq c_5^{N-n}\log^{N-n}(|D|/|d|),$$

with a constant c_5.

4. We now have to find a lower estimate for $\rho(k)$.

THEOREM 2. *Let $\epsilon > 0$ be a positive real number. There exists a constant $c_6(\epsilon)$ depending only on ϵ such that for all fields k normal over the field* P *of rational numbers the inequality holds*

$$\rho(k) \geqq c_6(\epsilon)^{-n}|d|^{-\epsilon}. \tag{18}$$

($n = [k:\mathrm{P}]$, d the discriminant of k).

Proof. We may assume $\epsilon \leqq 1/2$.

---

[12] Instead of using the number of zeros of $L(s,\psi_j)$ in the rectangle (15), we may also compare the functional equations of both sides of (11). It is also possible to obtain (16) and (17) without analytical tools. In the case of (16), one has to use the explicit value of the $\lambda(j)$ as given in I. The formula (17) can be deduced from Artin's conductor formula, *Journal für die reine und angewandte Mathematik*, vol. 164 (1931), pp. 1-11.

Case I.[13] *For all fields k normal over* P *the function $\zeta(s,k)$ does not vanish for real s with $1-\epsilon/n<s<1$.*

Then $\zeta(1-\epsilon/n,k)\leqq 0$ and a result of C. L. Siegel[14] (Lemma 1 of I) shows that

$$\rho(k)\geqq(1-\epsilon/n)(\epsilon/n)2^{-n}e^{-2\pi n}\,|\,d\,|^{-\epsilon/2n}.$$

It is evident that an inequality (18) holds in this case.

Case II. *There exists a normal field k_0 of degree n_0 over* P *such that $\zeta(s,k_0)=0$ for a real s with $1-\epsilon/n_0<s<1$.*

We may consider n_0, s, and the discriminant d_0 of k_0 as depending only on ϵ. Let K be the compositum kk_0 of the normal fields k and k_0. Then K is normal over k_0, and (11) for k_0 instead of k implies that $\zeta(s,K)=0$ for our fixed s, since the right hand side of (11) remains finite for this s. Again, Siegel's result[14] can be applied and yields

$$\rho(K)\geqq s(1-s)2^{-N}e^{-2\pi N}\,|\,D\,|^{-(1-s)/2}, \tag{19}$$

where $N=[K:\mathrm{P}]$ and where D is the discriminant of K;

$$N\leqq n_0n;\qquad |\,D\,|\leqq|\,d\,|^{n_0}\,|\,d_0\,|^{n}. \tag{20}$$

Hence

$$|\,D\,|^{-(1-s)/2}\geqq|\,D\,|^{-\epsilon/2n_0}\geqq|\,d\,|^{-\epsilon/2}\,|\,d_0\,|^{-n\epsilon/2n_0},$$

and since s, n_0, d_0 depend on ϵ but not on k, (19) can be written in the form

$$\rho(K)\geqq c_7(\epsilon)^{-n}\,|\,d\,|^{-\epsilon/2}, \tag{21}$$

where $c_7(\epsilon)$ depends only on ϵ.

On the other hand, K is normal over k and Theorem 1 can be applied. If we choose $\gamma=n_0/\epsilon$ and take (20) into account, we obtain an inequality

$$\rho(k)\geqq\rho(K)c_8(\epsilon)^{-n}\,|\,d\,|^{-\epsilon/2}, \tag{22}$$

where $c_8(\epsilon)$ also depends only on ϵ. Combination of (21) and (22) yields (18) with $c_6(\epsilon)=c_7(\epsilon)c_8(\epsilon)$, and Theorem 2 is established.

It follows from (18) and (3) that an inequality of the type

$$\log(Rh)-\log\sqrt{|d|}\geqq-nc_9(\epsilon)-2\epsilon\log\sqrt{|\,d\,|} \tag{23}$$

holds.

---

[13] In particular, we have Case I, if Riemann's conjecture is true for $\zeta(s,k)$. If $\zeta(s,k)$ does not vanish on the real line between $1-\epsilon/\log|\,d\,|$ and 1, the same method yields the better estimate $\rho(k)\geqq c(\epsilon)^{-n}(\log|\,d\,|)^{-1}$ with a positive $c(\epsilon)$ depending only on ϵ. This result is due to Hecke.

[14] *Loc. cit.*[3], Lemma 2 on p. 85.

Since $n/\log|d|$ is bounded for all algebraic number fields $k \neq \mathrm{P}$, we see that $\log(Rh)/\log\sqrt{|d|}$ is bounded below. This together with (6) yields:

THEOREM 3. *There exists an absolute constant* c_{10} *such that* $|\log(Rh)| \leqq c_{10}\log\sqrt{|d|}$ *for all fields* k *normal over* P.

Further, if k ranges over a sequence of fields normal over k for which $n/\log|d| \to 0$, then (23) implies that $\underline{\lim}\log(Rh)/\log\sqrt{|d|} \geqq 1 - 2\epsilon$. Since $\epsilon > 0$ is arbitrary, the limit inferior is at least 1. By (7), the corresponding limit superior is at most 1. Hence we have

THEOREM 4. *If* k *ranges over a sequence of fields normal over* P *for which* $n/\log|d| \to 0$, *then* $\log(Rh) \sim \log\sqrt{|d|}$.

5. We now consider arbitrary algebraic number fields k. If $k = \mathrm{P}(\alpha_1)$, and if $\alpha_1, \alpha_2, \cdots, \alpha_n$ are the conjugates of α_1, set

$$k^* = k(\alpha_1, \alpha_2, \cdots, \alpha_n).$$

Then k^* is the minimal normal field over P which contains k. The degree n^* of k^* over P satisfies the condition $n \leqq n^* \leqq n!$. We need an estimate for the discriminant d^* of k^* in terms of the discriminant d of k. Let θ_i denote the isomorphism of $k = \mathrm{P}(\alpha_1)$ on $\mathrm{P}(\alpha_i)$ which maps α_1 on α_i. Let $\mathfrak{e}_{ij}$ denote the greatest common divisor of all $\omega^{\theta_i} - \omega^{\theta_j}$ with integral $\omega \varepsilon k$; $(i, j$ fixed; $i < j)$. If we set $L_i = \mathrm{P}(\alpha_1, \alpha_2, \cdots, \alpha_i)$, $L_0 = \mathrm{P}$, the different of $L_i = L_{i-1}(\alpha_i)$ relative to L_{i-1} divides [15]

$$\prod_{j>i} \mathfrak{e}_{ij}$$

$(i = 1, 2, \cdots, n-1)$, and hence the different $\mathfrak{d}^*$of $k^* = L_{n-1}$ relative to P divides

$$\prod_{i=1}^{n-1} \prod_{j>i} \mathfrak{e}_{ij}.$$

However, the square of this double product is the discriminant d of k. Thus, $\mathfrak{d}^{*2} \mid d$. Taking the norm in k^* relative to P, we obtain $d^{*2} \mid d^{n^*}$ and this implies

$$|d^*| \leqq |d|^{n^*/2}. \tag{24}$$

Apply Theorem 1 to the field k and the normal extension field k^*. If $\gamma = 1/2\epsilon$, $\epsilon < 1/2$, we find the inequality

$$\rho(k) \geqq \rho(k^*) c_{11}(\epsilon)^{-n^*} |d^*|^{-\epsilon} \tag{25}$$

---

[15] See Hilbert, "Zahlbericht," *Jahresbericht der Deutschen Mathematiker-Vereinigung*, vol. 4 (1897), pp. 177-546; Chapter V.

with a suitable $c_{11}(\epsilon)$. On the other hand, Theorem 2 applied to k^* yields

$$(26) \qquad \rho(k^*) \geqq c_6(\epsilon)^{-n^*} \mid d^* \mid^{-\epsilon}.$$

On combining (3), (25), and (26), we obtain

$$(2\pi)^n Rh/\sqrt{\mid d \mid} \geqq \rho(k) \geqq (c_6(\epsilon) c_{11}(\epsilon))^{-n^*} \mid d^* \mid^{-2\epsilon}.$$

Hence, for a suitable $c_{12}(\epsilon)$,

$$\log (Rh) - \log\sqrt{\mid d \mid} \geqq - n^* c_{12}(\epsilon) - 2\epsilon \log \mid d^* \mid$$

and, by (24),

$$(27) \qquad \left(\frac{\log(Rh)}{\log\sqrt{\mid d \mid}} - 1\right)\frac{1}{n^*} \geqq - \frac{c_{12}(\epsilon)}{\log\sqrt{\mid d \mid}} - 2\epsilon.$$

Excluding a finite number of fields, we have $c_{12}(\epsilon)/\log\sqrt{\mid d \mid} \leqq \epsilon$. Thus, the left hand side of (27) is bounded below and does not possess any negative limit points, if we let k range over all algebraic number fields different from P. This, together with (8) yields

THEOREM 5. *If k ranges over all algebraic number fields different from* P *and if n^* denotes the degree over* P *of the minimal normal field k^* over* P *containing k, the set of all values*

$$\left(\frac{\log(Rh)}{\log\sqrt{\mid d \mid}} - 1\right)\frac{1}{n^*}$$

is bounded and possesses 0 as its only limit point.

UNIVERSITY OF MICHIGAN.

ON COMPLETENESS IN THE SENSE OF ARCHIMEDES.*

By L. W. COHEN and CASPER GOFFMAN.

An archimedean extension of an ordered abelian group G is a group $H \supset G$ such that if $h > \theta$, the identity, in H, then there are $g > \theta$ in G and an integer m such that $mg \geqq h$ and $mh \geqq g$. The group G is called complete in the sense of Archimedes or a-complete if it has no proper archimedean extension.

Topological or t-completeness may be defined for G in the manner of Cantor or, equivalently, in the manner of Dedekind [1, 2].[1] It has been shown [2] that if G is a-complete it is t-complete but that the converse is not true. The principal result of this paper is that G is a-complete if and only if, for every proper isolated subgroup $I \subset G$, the ordered factor group G/I is t-complete and non-discrete.

A group G is called non-discrete if to $g > \theta$ in G there is a $g' \varepsilon G$ such that $g > g' > \theta$. A subgroup $I \subset G$ is called isolated [3] if $|g| \leqq g' \varepsilon I$ implies $g \varepsilon I$. The subgroup consisting of θ alone is isolated and, in order to avoid confusion $I = [\theta]$ shall be referred to as degenerate when it occurs in the discussion. It is easy to see that if I is isolated the cosets $I_g = I + g = [g' + g \mid g' \varepsilon I]$ are segments of G and that the factor group G/I is ordered by the condtion $I_g \geqq I$, if $g \geqq \theta$ in G. The topology of a group G is determined by a limiting ordinal $\xi^* = \xi^*(G)$ with the property that if $\xi_\eta < \xi^*$ for $\eta < \eta^* < \xi^*$ then $\sup[\xi_\eta \mid \eta < \eta^*] < \xi^*$ and a strictly decreasing sequence $g_\xi \downarrow \theta$, $\xi < \xi^*$, in G. The neighborhoods in G are sets $U_\xi(g) = [g' \mid |g' - g| < g_\xi]$, $\xi < \xi^*$, $g \varepsilon G$.

LEMMA 1. *If G is non-discrete and t-complete, then the neighborhoods $U_\xi(\theta)$ of the identity are equivalent to* (a) *a sequence I_ξ, $\xi < \xi^*$, of isolated subgroups of G, or* (b) *the segments $S_n = [g \mid |g| < \gamma_n]$, $n < \omega$, of an isolated subgroup $I \subset G$ which is isomorphic to the additive group of real numbers.*

Proof. Consider the sequence $g_\xi \downarrow \theta$ in G. Then either (α) for each $\xi < \xi^*$ there is a g_{η_ξ} such that for all integers m, $mg_{\eta_\xi} < g_\xi$, or (β) there is a $\xi < \xi^*$ such that for all g_η, $\xi \leqq \eta < \xi^*$, there is an integer m_η such that $m_\eta g_\eta \geqq g_\xi$.

---

* Received September 17, 1949.

[1] Numbers in square brackets refer to the bibliography at the end of the paper.

Suppose (α) holds. Let I_ξ be the set of g such that for some $g_1, \cdots, g_n \,\varepsilon\, U_{\eta_\xi}(\theta)$ and integers $m_1, \cdots, m_n$, $|g| \leqq \sum_{i=1}^{n} m_i g_i$. Then $|g| \leqq g' \,\varepsilon\, I_\xi$ implies $g \,\varepsilon\, I_\xi$, and $g', g'' \,\varepsilon\, I_\xi$ implies $g' \pm g'' \,\varepsilon\, I_\xi$. Hence I_ξ is an isolated subgroup of G. Further, $U_{\eta_\xi}(\theta) \subset I_\xi \subset U_\xi(\theta)$, $\xi < \xi^*$. Since $g \,\varepsilon\, U_{\eta_\xi}(\theta)$ implies $|g| \,\varepsilon\, U_{\eta}\ (\theta)$, the first inclusion holds. Now suppose $g \,\varepsilon\, I_\xi$. Then $|g| \leqq \sum_{i=1}^{n} m_i g_i \leqq M g_0$, where $M = \sum_{i=1}^{n} |m_i|$, $g_0 = \max[|g_1|, \cdots, |g_n|] \,\varepsilon\, U_{\eta_\xi}(\theta)$. From ($\alpha$) it follows that $M g_0 \,\varepsilon\, U_\xi(\theta)$. Hence $g \,\varepsilon\, U_\xi(\theta)$ and the second inclusion follows. Hence (α) implies (a). Suppose (β) holds. Let $\gamma_0 = g_\xi$. Since $2g_\eta \downarrow \theta$, there is an ω-subsequence γ_n of g_η, $\xi \leqq \eta < \xi^*$, such that $2\gamma_n < \gamma_{n-1}$. Now ω-$\lim \gamma_n = \theta$. If not, there is a g_ζ, $\xi < \zeta < \xi^*$, such that $g_\zeta < \gamma_n$, $n < \omega$. Then for each m, $m g_\zeta < m\gamma_m < 2^m \gamma_m < \gamma_0 = g_\xi$. This contradicts ($\beta$). Hence $\xi^*(G) = \omega$ and $\lim \gamma_n = \theta$. From (β) it follows that for each $n < \omega$ there is an integer $m_n \geqq 1$ such that $m_n \gamma_n < \gamma_{n-1} < (m_n + 1)\gamma_n$. Now consider all g such that $\theta \leqq g < \gamma_0$. For each n there is a unique integer $j_n = j_n(g)$ such that $0 \leqq j_n \leqq m_n$ and $\sum_{i=1}^{n} j_i \gamma_i \leqq g < \gamma_n + \sum_{i=1}^{n} j_i \gamma_i$. Since $\theta \leqq g - \sum_{i=1}^{n} j_i \gamma_i < \gamma_n$, $\lim \sum_{i=1}^{n} j_i \gamma_i = g$. Now let $x_0 = 1$, $x_n = x_{n-1}/(m_n + 1)$, $n < \omega$. Then $\sum_{i=1}^{n} j_i x_i \leqq \sum_{i=1}^{n} x_{n-1} < 2$. It is easy to show that $x = x(g) = \sum_{i=1}^{\infty} j_i x_i$ is a homeomorphism between the set $[g \mid \theta \leqq g < g_\xi] \subset G$ and the set of real numbers $[x \mid \theta \leqq x < \sum_{i=1}^{\infty} x_{i-1}]$ since G is t-complete. Let I be the isolated subgroup generated by $[g \mid \theta \leqq g < g_\xi]$. The mapping $x = x(g)$ may be extended additively to I and this extension is an isomorphism between I and the additive group of real numbers. The segments $S_n = [g \mid |g| < \gamma_n]$ are equivalent to the neighborhoods of θ in G. Hence (β) implies (b).

LEMMA 2. *If G is a non-discrete group and H is an archimedean extension of G then H is non-discrete, $\xi^*(H) = \xi^*(G)$ and, if $V_\xi(h)$ is a neighborhood system for H, $U_\xi(g) = V_\xi(g) \cap G$ is a neighborhood system for G.*

Proof. H is non-discrete. If not, there is a smallest positive $h \,\varepsilon\, H$ and $h < g_\xi$, $\xi < \xi^*(G)$. Since H is an archimedean extension of G, there is a $g > \theta$ in G such that $g \leqq mh$ for some integer m. Then $\theta < g \leqq mh < mg_\xi$, $\xi < \xi^*(G)$ contrary to $\lim mg_\xi = \theta$ in G.

Now consider $h_\eta \downarrow 0$, $\eta < \xi^*(H)$, in H. For each h_η there is a $g_{\xi_\eta} \leqq h_\eta$. Because of the property of $\xi^*(G)$ mentioned above and the decreasing character of the g_ξ, the g_{ξ_η} can be chosen so that $\eta_1 < \eta_2 < \xi^*(H)$ implies

$g_{\xi_{\eta_2}} \leqq h_{\eta_2} < g_{\xi_{\eta_1}} \leqq h_{\eta_1}$. Hence $\xi^*(H)$ — $\lim g_{\xi_\eta} = \theta$ in G as well as in H. By the minimal property of $\xi^*(H)$ and $\xi^*(G)$ ([2], p. 311), $\xi^*(H) = \xi^*(G)$.

The remainder of the lemma follows at once.

LEMMA 3. *If G is a-complete and I is a proper isolated subgroup of G, then G/I is a-complete.*

Proof. Assume that G/I is not a-complete. Then there is a proper archimedean extension $K \supset G/I$. Let H be the set of pairs (k, j), $k \varepsilon K$, $j \varepsilon I$. We define addition by $(k_1, j_1) + (k_2, j_2) = (k_1 + k_2, j_1 + j_2)$. Since I is the identity in K, and θ is the identity in $I \subset G$, (I, θ) is the identity for this addition. We define order in H by the relation $(k, j) > (I, \theta)$ if $k > I$ in K or $k = I$ and $j > \theta$ in I. With these definitions, H is an ordered abelian group. The subgroup $H' = [(I_g, j) \mid g \varepsilon G, j \varepsilon I] \subset H$ is isomorphic with G. For: G being a-complete, $mx = g$ has a solution in G for all $g \varepsilon G$ and integers $m \neq 0$ ([4], p. 197). Since I is an isolated subgroup of G, $mI = I$ for all integers $m \neq 0$. From a known theorem[2] ([5], p. 766) it follows that I is a summand of G. Hence $H' \sim G$.

There is a $k_0 \varepsilon K - G/I$ and so the pair $(k_0, \theta) \varepsilon H - H'$. Now consider $(k, j) > (I, \theta)$ in H. If $k = I$, then $(k, j) \varepsilon H' \sim G$. If $k > I$, then since K is an archimedean extension of G/I, there is an $I_g > I$ such that for some integer m, $mI_g > k$ and $mk > I_g$. Now $m(I_g, j) = (mI_g, mj) > (k, j)$ and $m(k, j) = (mk, mj) > (I_g, j)$. Hence H is a proper archimedean extension of $H' \sim G$, contrary to the hypothesis.

LEMMA 4. *If G is a-complete then G is non-discrete.*

Proof. If $g > 0$ in G then $g > h > 0$ where h is the solution of $2x = g$.

We introduce some notation for convenience. Consider an extension H of an ordered abelian group G and an isolated subgroup $I \subset G$. We denote by I' the set of $h \varepsilon H$ for which there are $g(h) \varepsilon I$ such that $|h| \leqq g(h)$. It is clear that I' is an isolated subgroup of H. The mapping $f(I_g) = I'_g$ is an order preserving isomorphism of G/I into H/I'.

LEMMA 5. *If G is a non-discrete and t-complete group and H is an archimedean extension of G, then for $h \varepsilon H - G$ there is an isolated non-degenerate subgroup $I \subset G$ such that $I'_h \cap G = 0$.*

Proof. Consider $h \varepsilon H - G$. Since G is t-complete and non-discrete, there is a neighborhood $V_\xi(h)$ such that $V_\xi(h) \cap G = 0$. G satisfies the hypothesis of Lemma 1. If (a) of Lemma 1 holds for G, then by Lemma 2, $U_\xi(\theta) = V_\xi(\theta) \cap G$ is a neighborhood of θ in G, and it contains an isolated subgroup $I_\xi \subset G$. Now $I'_\xi \subset V_\xi(\theta)$, and $(I'_\xi + h) \cap G \subset V_\xi(h) \cap G = 0$. Now suppose (b) of Lemma 1 holds for G. There is an isolated subgroup $I \subset G$, isomorphic to the additive group of real numbers, such that $U_\xi(\theta) \subset I$ and is mapped on some set of real nuumbers $[x \mid |x| < \epsilon]$, $\epsilon > 0$. Since the additive group of real numbers has no proper archimedean extension, $I' = I$. Now $(I' + h) \cap G = (I + h) \cap G = 0$ since $h \notin G$.

---

[2] The authors are indebted to the referee for calling this theorem to their attention.

Suppose that H is an archimedean extension of G and that σ is the set of all isolated subgroups $I \subset G$ such that $(I' + h) \cap G = 0$. The union $K = \bigcup_\sigma I$ is an isolated subgroup of G such that $(K' + h) \cap G = 0$. We shall call K the maximal isolated subgroup with this property.

THEOREM. *An ordered abelian group G is a-complete if and only if, for every isolated proper subgroup $I \subset G$, G/I is non-discrete and t-complete.*

Proof. Suppose that the condition holds and that H is an archimedean extension of G. We show that $H = G$. If not, there is an $h_0 \varepsilon H - G$. From Lemma 5, and the above remark, there is a maximal isolated subgroup $K \subset G$ such that $(K' + h_0) \cap G = 0$ and K is not degenerate. Now $K \neq G$. If not, then since there is in $K = G$ an element $g > |h_0|$, $K' + h_0 = K' \supset K$. This contradicts $(K' + h_0) \cap G = 0$.

Now H/K' is an archimedean extension of $f(G/K)$ where $K'_g = f(K_g)$ is the isomorphism of G/K into H/K' defined previously. For: if $K'_h > K'$ in H/K', then $h > \theta$ in H. Since H is an archimedean extension of G, there is a $g > 0$ such that for some integer m, $mg > h$ and $mh > g$. Hence $f(K_{mg}) = K'_{mg} \geqq K'_h$ and $K'_{mh} \geqq K'_g = f(K_g)$. Further, $K' + h_0 \varepsilon H/K' - f(G/K)$. If not, there is $K'_g = K' + h_0 \varepsilon f(G/K)$, and so $(K' + h_0) \cap G \supset K_g \neq 0$, which is false.

Now by hypothesis G/K is non-discrete and t-complete so that Lemma 5 applies to $H/K' \supset f(G/K)$ and $K'_{h_0} \varepsilon H/K - f(G/K)$. Hence there is an isolated subgroup $\mathfrak{J} \subset f(G/K)$ such that $(\mathfrak{J}' + K') \cap f(G/K) = 0$.

If $K'_g \varepsilon \mathfrak{J}'$ then $g \varepsilon K$. If not, there is $g_0 \varepsilon G - K$ such that $K' + g_0 \varepsilon \mathfrak{J}'$. Let J be the set of $g \varepsilon G$ such that for some $g' \varepsilon K$ and integer m, $|g| \leqq g' + mg_0$. Then J is an isolated subgroup of G and $K \subset J - [g_0]$. Since K is the maximal isolated subgroup of G such that $(K' + h_0) \cap G = 0$, there is a $g_1 \varepsilon (J' + h_0) \cap G$. Now $g_1 - h_0 \varepsilon J'$, and so there are $g \varepsilon J$ and $g' \varepsilon K$ and an integer m, such that $|g_1 - h_0| \leqq |g| \leqq g' + mg_0$. Hence $K'_{|g_1 - h_0|} \leqq K'_{g'} + K'_{mg_0} = K'_{mg_0} = mK'_{g_0} \varepsilon \mathfrak{J}'$; since $\mathfrak{J}'$ is an isolated subgroup of H/K', and $K'_{g'} = K'$ is the identity in $\mathfrak{J}'$. Further, $K_{g_1 - h_0} \varepsilon \mathfrak{J}'$, $K'_{g_1} \varepsilon \mathfrak{J}' + K'_{h_0}$ and so $K'_{g_1} \notin f(G/K)$. Hence $K_{g_1} \notin G/K$, contrary to $g_1 \varepsilon G$.

Since $\mathfrak{J} \neq [K]$ by Lemma 5, $\mathfrak{J}' \neq [K']$. Hence there is $h > \theta$ in H such that $K' < K'_h \varepsilon \mathfrak{J}'$. Consider any $g > \theta$ in G. If $g \varepsilon K$, then for all integers m, $mg \varepsilon K \subset K' < K'_h$. Hence $mg < h$ for all integers m. If $g \not\varepsilon K$, $K_g \subset K'_g \not\varepsilon \mathfrak{J}'$. Since for all integers m, $K'_{mh} = mK'_h \varepsilon \mathfrak{J}'$, it follows that if $g > \theta$, $K'_{mh} < K'_g$. Hence $mh < g$ for all integers m. Thus there does not correspond to some $h > \theta$ in H any $g > \theta$ in G such that, for some integer m, $mg \geqq h$ and $mh \geqq g$. This contradicts the assumption that H is an archimedean extension of G. Hence $H = G$, and the sufficiency is proved.

To establish the necessity of the condition, consider an a-complete group G. If I is a proper isolated subgroup of G, then G/I is an ordered abelian group. By Lemma 3, G/I is a-complete. Since an a-complete ordered abelian group is t-complete ([2], Theorem 9), G/I is t-complete. By Lemma 4, the a-complete group G/I is non-discrete.

We conclude with examples indicating the independence of the conditions. Let R be the group of rational valued sequences $r = \{r_n\}$ ordered lexicographically. For neighborhoods of the identity we use $U_n(\theta) = [r \mid r_1 = \cdots = r_n = 0]$. The $U_n(\theta)$ are isolated subgroups of R. R is non-discrete and t-complete, but the factor groups $R/U_n(\theta)$ are not t-complete, although they are non-discrete. R is not a-complete since R^*, the group of all real sequences $x = \{x_n\}$, is a proper archimedean extension. Let N be the group of complex numbers $n + iy$, where n is an integer and y is real, ordered lexicographically. The neighborhoods of the identity are $U_n(\theta) = [iy \mid |y| < 1/n]$. N is non-discrete and t-complete. The subgroup $I = [iy \mid y \text{ real}]$ is isolated. The factor group N/I, isomorphic to the group of real integers, is discrete and, therefore, trivially t-complete. N is not a-complete, since the ordered group C of all complex numbers is a proper archimedean extension of N.

QUEENS COLLEGE, FLUSHING, N. Y.
UNIVERSITY OF OKLAHOMA, NORMAN, OKLA.

BIBLIOGRAPHY.

[1] R. Baer, "Zur Topologie der Gruppen," *Jour. für d. reine u. angew. Math.*, vol. 160 (1929), pp. 208-226.

[2] L. W. Cohen and C. Goffman, "The topology of ordered abelian groups," *Transactions of the American Mathematical Society*, vol. 67 (1994), pp. 310-319.

[3] W. Krull, "Allgemeine Bewertungstheorie," *Jour. für d. reine u. angew. Math.*, vol. 167 (1932), pp. 160-196.

[4] F. Hausdorff, *Grundzüge der Mengenlehre*, Leipzig (1914).

[5] R. Baer, "The subgroup of elements of finite order of an abelian group," *Annals of Mathematics*, vol. 37 (1936), pp. 766-781.

ON COMPLETENESS AND CATEGORY IN UNIFORM SPACE.*

By L. W. Cohen and Casper Goffman.

The problem of the relation between completeness and category in uniform space arises from results obtained for certain topological groups. For ordered abelian groups G, two concepts of completeness have been defined: G is called t-complete if every fundamental sequence has a limit in G; G is called a-complete if it has no proper archimedean extension [1, 2].[1] We have shown [1] that if G is a-complete, it is t-complete and of the second ξ^*-category.[2] There are t-complete G which are not a-complete, but it is not known whether there are t-complete G which are of the first ξ^*-category. However there are, for each $\xi^* > \omega$, topological abelian groups which are t-complete and of the first ξ^*-category [2]. The relation between the two concepts of completeness for ordered abelian groups G is given in the theorem [3]: G is a-complete if and only if, for every isolated subgroup $I \subset G$, the factor group G/I is non-discrete and t-complete.

The property of Archimedes, requiring order for its formulation, is not applicable to topological groups in general. Thus a property corresponding to a-completeness which would imply that such a group is of the second ξ^*-category is lacking. It is the purpose of this paper to supply such a concept in a form applicable to the class of uniform spaces metrizable with distances in an ordered abelian group.

We consider the uniform spaces $S = [x]$ with neighborhood systems $\mathfrak{U} = [U_\xi(x), x \in S, \xi < \xi^*]$ satisfying the conditions:

(1) $\bigcap\limits_{\xi < \xi^*} U_\xi(x) = (x)$;

(2) $\xi_1 < \xi_2 < \xi^*$ implies $U_{\xi_1}(x) \supset U_{\xi_2}(x)$;

(3) $\eta < \xi^*$ implies there is a $\xi(\eta)$ such that $\eta < \eta^* < \xi^*$, and if $U_{\xi(\eta)}(y) \cap U_{\xi(\eta)}(x) \neq 0$ then $U_{\xi(\eta)}(y) \subset U_\eta(x)$;

---

* Received November 7, 1949.

[1] Numbers in brackets refer to the appended bibliography.

[2] There is a unique ordinal ξ^* associated with an ordered abelian group G. A sequence in G is a single valued function g_ξ on $\xi < \xi^*$ to G. A group G is of the first or second ξ^*-category according as it is or is not the union of nowhere dense sets N_ξ, $\xi < \xi^*$.

(4′) $\eta^* < \xi^*$ implies $\bigcap_{\eta<\eta^*} U_{\xi_\eta}(x)$ is open, where ξ^* is a limiting ordinal such that if $\xi_\eta < \xi^*$ on $\eta < \eta^* < \xi^*$, then

$$(*) \qquad \sup\,[\xi_\eta \mid \eta < \eta^*] < \xi^*.$$

These conditions are satisfied by ordered abelian groups G with segments as neighborhoods. It has been shown [2] that if S is t-complete and satisfies

(4) If $U_{\xi_\eta}(x_\eta)$ decreases on $\eta < \eta^* < \xi^*$ then $\bigcap_{\eta<\eta^*} U_{\xi_\eta}(x_\eta)$ is a non-empty open set,

then S is of the second ξ^*-category.

We shall associate with S a family of "factor" spaces S^λ, analogous to the factor groups G/I mentioned above, and call S a-complete if, for all λ, S^λ is t-complete. It will be shown that if S is a-complete then S is of the second ξ^*-category. It has been shown [4] that a uniform space S is metrizable with distances in an ordered abelian group G if and only if S satisfies (1), (2), (3), (4′). Our result may now be stated as follows: A metrizable uniform space S which is a-complete is of the second ξ^*-category. Since the two concepts of completeness coincide if $\xi^* = \omega$, it would seem that the proper generalization of the classical concept of completeness is a-completeness.

We assume $\xi^* > \omega$ in the remainder of this note. In [4] it is shown that S has a neighborhood system $\mathcal{V} = [V_\xi(x)]$, equivalent to $\mathcal{U}$, which satisfies (1), (2), (3), (4′) and

(5) $V_\xi(x) \cap V_\xi(y) \neq 0$ implies $V_\xi(x) = V_\xi(y)$.

We define a neighborhood system $\mathcal{W} = [W_\xi(x)]$ where $W_\xi(x) = \bigcap_{\eta<\xi} V_\eta(x)$, $\xi < \xi^*$. It is clear that $\mathcal{W}$ is equivalent to $\mathcal{V}$ and satisfies (1), (2), (3), (4′), (5) and

(6) If $\lambda < \xi^*$ is a limiting ordinal, then $W_\lambda(x) = \bigcap_{\xi<\lambda} W_\xi(x)$.

LEMMA 1. *If $\xi < \eta < \xi^*$, then $W_\xi(y) = \bigcup\,[W_\eta(x) \mid x \in W_\xi(y)]$.*

Proof. Consider $z \in W_\eta(x)$ where $x \in W_\xi(y)$. Then $z \in W_\eta(x) \subset W_\xi(x) = W_\xi(y)$ by (2) and (5). Hence $\bigcup\,[W_\eta(x) \mid x \in W_\xi(y)] \subset W_\xi(y)$. The reversed inclusion follows from (1).

We consider any limiting ordinal $\lambda \leqq \xi^*$. Let Λ be the set of limiting ordinals $\mu \leqq \lambda$ such that there is a strictly increasing, ordinal valued function $\xi(\nu)$ on $\nu < \mu$ for which $\sup[\xi(\nu) \mid \nu < \mu] = \lambda$. Since $\lambda \in \Lambda$, Λ is not empty and so contains a first ordinal $\mu^*(\lambda)$.

LEMMA 2. *$\mu^*(\lambda)$ has property* $(^*)$.

Proof. Suppose $\mu^*(\lambda)$ does not have property $(^*)$. Then there is a $\nu^* < \mu^*(\lambda)$ and an ordinal valued function $\eta(\nu)$ on $\nu < \nu^*$ such that $\sup[\eta(\nu) \mid \nu < \nu^*] = \mu^*(\lambda)$. We define $\zeta(1) = \nu(1)$ and assume that, for $\sigma < \nu^*$, there is a strictly increasing, ordinal valued function $\zeta(\tau)$ on $\tau < \sigma$ such that $\eta(\tau) \leqq \zeta(\tau) < \mu^*(\lambda)$. Let $\zeta_0 = \sup[\zeta(\tau) \mid \tau < \sigma]$. If $\zeta_0 = \mu^*(\lambda)$, then σ is a limiting ordinal $\sigma^* < \nu^* < \mu^*(\lambda)$. If $\zeta_0 < \mu^*(\lambda)$, then we define $\zeta(\sigma) = \max[\eta(\sigma), \zeta_0] + 1$. In any case, there is a limiting ordinal $\sigma^* \leqq \nu^* < \mu^*(\lambda)$ and a strictly increasing function $\zeta(\sigma)$ on $\sigma < \sigma^*$ such that $\sup[\zeta(\sigma) \mid \sigma < \sigma^*] = \mu^*(\lambda)$.

Since $\mu^*(\lambda) \varepsilon \Lambda$, there is a strictly increasing function $\xi(\nu)$ on $\nu < \mu^*(\lambda)$ such that $\sup [\xi(\nu) \mid \nu < \mu^*(\lambda)] = \lambda$. Hence $\sup [\xi(\zeta(\sigma)) \mid \sigma < \sigma^*] = \lambda$. Since $\xi(\zeta(\sigma))$ is strictly increasing on $\sigma < \sigma^*$, $\sigma^* \varepsilon \Lambda$. Hence $\mu^*(\lambda) \leqq \sigma^*$, contrary to $\sigma^* \leqq \nu^* < \mu^*(\lambda)$, and the lemma is proved.

Now for each limiting ordinal $\lambda < \xi^*$ we define $X^\lambda = W_\lambda(x)$. We define S^λ as the set of all X^λ. For $\mu^*(\lambda)$ we fix an ordinal valued, strictly increasing function ξ_η on $\eta < \mu^*(\lambda)$ for which $\sup [\xi_\eta \mid \eta < \mu^*(\lambda)] = \lambda$. We define $W_\eta{}^\lambda(X^\lambda)$ as the set of $Y^\lambda = W_\lambda(y) \subset W_{\xi_\eta}(x)$ where $X^\lambda = W_\lambda(x)$. Finally we define for S^λ the neighborhood system $W^\lambda = [W_\eta{}^\lambda(X^\lambda), X^\lambda \varepsilon S^\lambda, \eta < \mu^*(\lambda)]$. It is consistent with these definitions to put $S^{\xi^*} = S$, $X^{\xi^*} = x$, $W_\xi{}^{\xi^*}(X^{\xi^*}) = W_\xi(x)$.

LEMMA 3. *$S^\lambda, \lambda < \xi^*$, satisfies* (1), (2), (5) *with ξ^* replaced by $\mu^*(\lambda)$*.

Proof. Condition (1) follows from (6). Condition (2) and (5) are clearly satisfied.

The concepts of fundamental sequence and t-completeness are defined for S^λ in terms of $\mu^*(\lambda)$ as in [2]. We call S a-complete (complete in the sense of Archimedes) if S^λ is t-complete for each $\lambda \leqq \xi^*$. Since $S^{\xi^*} = S$, it follows that if S is a-complete it is t-complete.

LEMMA 4. *If S is a-complete then S satisfies* (4).

Proof. Consider neighborhoods $W_{\alpha_\beta}(x_\beta)$ which decrease on $\beta < \beta^* < \xi^*$. By property $(^*)$, $\sup [\alpha_\beta \mid \beta < \beta^*] = \xi < \xi^*$. If there is a $\gamma < \beta^*$ such that $\alpha_\gamma = \xi$, $\bigcap_{\beta < \beta^*} W_{\alpha_\beta}(x_\beta) = W_{\alpha_\gamma}(x_\gamma)$ by (2), and is open by (4').

If there is no such γ, then ξ is a limiting ordinal $\lambda < \xi^*$. We now consider the space S^λ, the ordinal $\mu^*(\lambda)$ and the fixed function ξ_η such that $\sup [\xi_\eta \mid \eta < \mu^*(\lambda)] = \lambda$.

There exists a function α_{β_η} strictly increasing on $\eta < \mu^*(\lambda)$ such that $\sup [\alpha_{\beta_\eta} \mid \eta < \mu^*(\lambda)] = \lambda$. For there is a β_1 such that $\xi_1 < \alpha_{\beta_1}$. Assume that α_{β_η} is strictly increasing on $\eta < \sigma < \mu^*(\lambda)$ and that $\xi_\eta \leqq \alpha_{\beta_\eta}$. Then since $\mu^*(\lambda)$ is the smallest ordinal in Λ, $\sup [\alpha_{\beta_\eta} \mid \eta < \sigma] = \alpha < \lambda$. Now $\max [\alpha, \xi_\sigma] < \lambda$, and so there is a $\beta_\sigma < \beta^*$ such that $\max [\alpha, \xi_\sigma] < \alpha_{\beta_\sigma} < \lambda$. Transfinite induction yields α_{β_η} strictly increasing on $\eta < \mu^*(\lambda)$ for which $\sup [\alpha_{\beta_\eta} \mid \eta < \mu^*(\lambda)] = \lambda$.

To simplify notation let $\alpha_{\beta_\eta} = \delta_\eta$ and $x_{\beta_\eta} = y_\eta$. We define $X^\lambda{}_\eta = W_\lambda(y_\eta) \varepsilon S^\lambda$, $\eta < \mu^*(\lambda)$, and show that it is a fundamental sequence. Consider $\zeta \leqq \eta < \mu^*(\lambda)$. Then $\xi_\zeta \leqq \delta_\zeta \leqq \delta_\eta < \lambda$, and $X^\lambda{}_\eta = W_\lambda(y_\eta) \subset W_{\delta_\eta}(y_\eta) \subset W_{\delta_\zeta}(y_\zeta)$ by (2) and the decreasing character of the neighborhoods $W_{\alpha_\beta}(x_\beta)$. It follows that $X^\lambda{}_\eta \varepsilon W^\lambda{}_\zeta (X_\zeta)$ for $\zeta \leqq \eta < \mu^*(\lambda)$, and so $X^\lambda{}_\eta$ is a fundamental sequence in S^λ.

Since S^λ is t-complete, $\lim X^\lambda{}_\eta = X^\lambda \varepsilon S^\lambda$. There is an $x \varepsilon S$ such that $X^\lambda = W_\lambda(x)$. We show that $W_\lambda(x) \subset \bigcap_{\beta<\beta^*} W^\lambda{}_{\alpha\beta}(x_\beta)$. If not, there is a $\beta < \beta^*$ such that $W_\lambda(x) \not\subset W_{\alpha_\beta}(x_\beta)$. There is a $\zeta < \mu^*(\lambda)$ such that $\alpha_\beta < \delta_\zeta < \lambda$. There is a $\rho < \mu^*(\lambda)$ such that $\delta_\zeta < \xi_\rho < \lambda$. It follows from Lemma 1 that $W_{\delta_\zeta}(y_\zeta) = \bigcup [W_{\xi_\rho}(y) \mid y \varepsilon W_{\delta_\zeta}(y_\zeta)]$. Hence either $W_{\xi_\rho}(x) \subset W_{\delta_\zeta}(y_\zeta)$ or $x \not\varepsilon W_{\delta_\zeta}(y_\zeta)$. Now, by (2), $W_\lambda(x) \subset W_{\xi_\rho}(x) \subset W_{\delta_\zeta}(y_\zeta) \subset W_{\alpha_\beta}(x_\beta)$ follows from the first alternative, contrary to $W_\lambda(x) \not\subset W_{\alpha_\beta}(x_\beta)$. Hence $x \not\varepsilon W_{\delta_\zeta}(y_\zeta)$. From (5) it follows that $W_{\delta_\zeta}(x) \cap W_{\delta_\zeta}(y_\zeta) = 0$. Recalling that $\delta_\zeta < \xi_\rho < \lambda$, we have, for $\zeta \leqq \eta < \mu^*(\lambda)$, $W_{\xi_\rho}(x) \cap W_\lambda(y_\eta) \subset W_{\delta_\zeta}(x) \cap W_{\delta_\eta}(y_\eta) \subset W_{\delta_\zeta}(x) \cap W_{\delta_\zeta}(y_\zeta) = 0$ by (2), the increasing character of $\delta_\eta = \alpha_{\beta_\eta}$ on $\eta < \mu^*(\lambda)$ and the decreasing character of $W_{\alpha_\beta}(x_\beta)$ on $\beta < \beta^*$. Finally, from $W_{\xi_\rho}(x) \cap W_\lambda(y_\eta) = 0$ we have $X^\lambda{}_\eta \not\varepsilon W^\lambda{}_\rho(X^\lambda), \zeta \leqq \eta < \mu^*(\lambda)$. This contradicts $\lim X^\lambda{}_\eta = X_\lambda$. Hence $\bigcap_{\beta<\beta^*} W_{\alpha_\beta}(x_\beta) \neq 0$.

It remains to show that $\bigcap_{\beta<\beta^*} W_{\alpha_\beta}(x_\beta)$ is open. Consider $y \varepsilon \bigcap_{\beta<\beta^*} W_{\alpha_\beta}(x_\beta)$. From (4′) it follows that $W_{\alpha_\beta}(x_\beta)$, $\beta < \beta^*$, is open and so there is a ξ_β, $\beta < \beta^*$, such that $W_{\xi_\beta}(y) \subset W_{\alpha_\beta}(x_\beta)$. Again, by (4′), there is a $W_\xi(y) \subset \bigcap_{\beta<\beta^*} W_{\xi_\beta}(y) \subset \bigcap_{\beta<\beta^*} W_{\alpha_\beta}(x_\beta)$, and the set is open.

Since it has been shown that $\bigcap_{\beta<\beta^*} W_{\alpha_\beta}(x_\beta)$ is a non-empty open set it follows that S satisfies condition (4).

THEOREM. *If S is a-complete, S is of the second ξ^*-category.*

Proof. Since S is a-complete, it is t-complete. By Lemma 4, S satisfies (4). By Theorem 7 of [2] it follows that S, satisfying (1), (2), (3), (4) and being t-complete, is of the second ξ^*-category.

This theorem may be put in the following form.

THEOREM. *A uniform space S, metrizable with distances in an ordered abelian group, which is a-complete, is of the second ξ^*-category.*

Proof. It is shown [4] that S satisfies (1), (2), (3), (4′). The conclusion follows from the previous theorem.

QUEENS COLLEGE, FLUSHING, N. Y.
UNIVERSITY OF OKLAHOMA, NORMAN, OKLA.

BIBLIOGRAPHY.

[1] L. W. Cohen and C. Goffman, "The topology of ordered abelian groups," *Transactions of the American Mathematical Society*, vol. 67 (1949), pp. 310-319.

[2] ———, and C. Goffman, "A theory of transfinite convergence," *ibid.*, vol. 66 (1949), pp. 65-74.

[3] ———, and C. Goffman, "On completeness in the sense of Archimedes," *American Journal of Mathematics*, vol. 72 (1950), pp. 747-751.

[4] ———, and C. Goffman, "On the metrization of uniform space," *Proceedings of the American Mathematical Society*, to appear.

[5] M. Fréchet, "La notion d'uniformité et les écarts abstraits," *Comptes Rendus, Academie des Sciences*, vol. 221 (1945), p. 337.[*]

[6] R. Doss, "Sur la théorie de l'écart abstrait de M. Fréchet," *Bulletin des Sciences Mathématiques*, ser. 2, vol. 71 (1947), pp. 110-122.[*]

---

[*] We are indebted to the referee for calling our attention to [5, 6]. The common elements in the structure of the Fréchet space [6, p. 117] and the space of this paper permit us to state the following theorem: The space of Fréchet [5] is metrizable with distances in an ordered abelian group.

ON THE FUNDAMENTAL EQUATIONS OF DIFFERENTIAL GEOMETRY.*

By Philip Hartman and Aurel Wintner.

Introduction.

If a binary, symmetric, positive-definite matrix (g_{ik}) of functions $g_{ik} = g_{ik}(u_1, u_2)$ is given on a (u_1, u_2)-domain, then the problem of embedding $ds^2 = g_{ik}\, du_i\, du_k$ into a Euclidean (x_1, x_2, x_3)-space requires the determination of three functions $x_a = x_a(u_1, u_2)$ corresponding to which $\Sigma: x_a = x_a(u_1, u_2)$ is a surface on which $dx_a dx_a$ becomes identical with ds^2. Clearly, this problem presents a system of three non-linear partial differential equations for the three unknown functions $x_a(u_1, u_2)$. But it has never been attacked directly. In fact, differential geometry considers, besides the given form $g_{ik}\, du_i\, du_k$, an auxiliary quadratic form, $h_{ik}\, du_i\, du_k$, which is required to become the second fundamental form of the unknown surface Σ. What is then required is that the three unknown functions h_{ik} should satisfy a system of three partial differential equations which contain the three given functions g_{ik}, and which consist of the Theorema Egregium and of the two equations of Mainardi-Codazzi. This system of $1 + 2$ equations, usually considered under the analyticity assumption of Cauchy-Kowalewski, was treated in [4], along with the remaining issue of the actual embedding, with such a minimum of "real" assumptions of mere differentiability as is attainable today.

In Part 1 of the present paper the "remaining issue of embedding" will be treated under assumptions which are less generous and more natural than those which have to be assumed in the classical treatment of this problem. The latter consists of solving, with given g_{ik} and given h_{ik}, the so-called derivation equations of surface theory. These are supplied by the derivation formulae of Gauss and of Weingarten; derivation formulae the system of which is a "total" system, of the Jacobi type, for the Cartesian coordinates $x_a(u_1, u_2)$ of Σ, with the $1 + 2$ partial differential equations of the Theorema Egregium and of Mainardi-Codazzi as integrability conditions.

It was pointed out in [4], §2, that the $1 + 2$ partial differential equations contain a certain type of paradox in differentiability for the three unknown functions $x_a(u_1, u_2)$, if the given functions $g_{ik}(u_1, u_2)$, $h_{ik}(u_1, u_2)$ are referred to an actual Σ of class C^n, where $n < \infty$. A corresponding paradox presents

---

* Received January 24, 1950.

itself in the usual treatment of the derivation formulae. This will be disposed of in Part 1, by lightening the classical assumptions of differentiability. The latter are necessitated by the usual formulation of the $1+2$ equations which represent conditions of integrability for the total system.

Accordingly, what is involved is a formulation of the integrability conditions of a total system in such a way that no superfluous assumptions of differentiability are made. This will be accomplished by using integral conditions (along arbitrary closed curves of integration), instead of the local field conditions of the standard theory of total systems.

The total system in the differential geometry of surfaces is a linear one. It will, however, be seen at the end of Part 1 that a refined formulation of the existence theorem holds for non-linear total systems also. The reduction of the condition of integrability is relevant for other applications also; cf. e. g., [8].

Part 3 has about the same bearing on the derivation formulae of the differential geometry of curves (Frenet) as Part 1 does on those of surfaces (Bonnet). It is true that the differential equations are now ordinary, and therefore nothing like the existence theorem of Part 1 is involved. Nevertheless, the usual treatment contains again a "paradox in differentiability." The latter is due to the standard definition of torsion; it will be dissolved by a more geometrical approach.

Part 2 is, in the main, independent of the rest of the paper. It deals, on the C^2-level, with the various definitions of a torse and their equivalence.

Part 1. The Fundamental Equations of Surfaces.

1. Let X denote the vector (x, y, z). The vector function $X(u, v)$, defined on a (u, v)-domain, is called a surface of class C^n, where $n \geqq 1$, if each component of X is a function of class C^n and, in addition, the vector product (X_1, X_2) does not vanish (the subscripts 1, 2 denote partial differentiation with respect to u, v, respectively). Let

$$(1) \qquad N = (X_1, X_2)/|(X_1, X_2)|$$

be the unit normal vector of the surface.

On a surface of class C^2, the derivation formulae of Gauss,

$$(2) \qquad dX_i = (\Gamma^{\alpha}{}_{i\beta} X_\alpha + h_{i\beta} N)\, du^\beta \text{ for } i = 1, 2,$$

and those of Weingarten,

$$(3) \qquad dN = -\, g^{\alpha\beta} h_{\gamma\alpha} X_\beta du^\gamma,$$

must hold, if (u^1, u^2) denotes (u, v), and (g_{ij}), (h_{ij}) are the matrices of the first and second fundamental forms of the surface, (g^{ij}) is the reciprocal matrix $(g_{ij})^{-1}$, finally the $\Gamma^i{}_{jk}$ are the Christoffel symbols of the second kind.

If X is a surface of class C^3, the integrability conditions of (2) and (3), which state that $N_{12} = N_{12}$ and that $X_{i12} = X_{i21}$ for $i = 1, 2$, are equivalent to the equations of Mainardi-Codazzi,

$$(4) \qquad \partial h_{i1}/\partial v - \partial h_{i2}/\partial u = \Gamma^\alpha{}_{i2} h_{\alpha 1} - \Gamma^\alpha{}_{i1} h_{\alpha 2} \text{ for } i = 1, 2,$$

and to the Theorema Egregium,

$$(5) \qquad [(g_{11v} - g_{12u})/g]_v - [(g_{12v} - g_{22u})/g]_u = -(2K + \tfrac{1}{2}D/g^4)g,$$

where g denotes $(\det g_{ik})^{\frac{1}{2}}$, and K the representation $\det h_{ik}/\det g_{ik}$ of the Gaussian curvature, finally D is a 3-rowed determinant in which the first column consists of g_{11}, g_{12}, g_{22}, and the second and third columns are formed by partial derivatives of g_{11}, g_{12}, g_{22} with respect to u and v, respectively.

Conversely, the classical existence theorem of Bonnet states that if the symmetric matrices $(g_{ik}(u, v))$ and $(h_{ik}(u, v))$ are of class C^2 and class C^1, respectively, and if (g_{ik}) is positive-definite, finally if the integrability conditions, (4) and (5), are satisfied, then there exist surfaces $X(u, v)$ of class C^3 for which (g_{ik}), and (h_{ik}) become the matrices of the first and second fundamental forms, respectively.

2. There is an unsatisfactory element in this theorem. First, since every surface of class C^2 (not necessarily C^3) possesses first and second fundamental forms, whose matrices (g_{ik}), (h_{ik}) are of class C^1 and C^0, respectively, it is meaningful to ask whether two given symmetric matrices, (g_{ik}) and (h_{ik}), of class C^1 and C^0, respectively, are the matrices of the first and second fundamental forms of a surface X of class C^2. But Bonnet's theorem furnishes an answer to this question only if further smoothness assumptions on the g_{ik} and h_{ik} are made; and correspondingly, in the affirmative cases, it assures the existence of surfaces X which are smoother than required by this question.

It turns out that these objections can be overcome by restating the equations (4) and (5) in an "integrated" form. The equations (4) and (5) are thus replaced by the requirement that the relations

$$(6) \qquad \int_J h_{i1} du + h_{i2} dv = \int\int_S (\Gamma^\alpha{}_{i1} h_{\alpha 2} - \Gamma^\alpha{}_{i2} h_{\alpha 1})\, du\, dv, \text{ where } i = 1, 2,$$

and

$$(7) \qquad \int_J g^{-1}\{(g_{11v} - g_{12u}) du + (g_{12v} - g_{22u}) dv\} = \int\int_S (2K + \tfrac{1}{2}Dg^{-4}) g\, du\, dv$$

should hold whenever J is a Jordan curve of class C^1, and S the interior of J.

Whereas (4) and (5) need not be meaningful on a surface of class C^2, the relations (6) and (7) are meaningful and hold on every surface of class C^2. This follows from the considerations below. Incidentally, (7) shows that the Theorema Egregium, which is usually stated for surfaces of class C^3, is true for surfaces of class C^2. Actually, this statement can be deduced from the fact that the theorem of Gauss-Bonnet holds on surfaces of class C^2 (Weyl [9], pp. 42-44; cf. van Kampen, [5], p. 136).

The modified version of the theorem of Bonnet, which meets the objections mentioned above, is as follows:

(I) *Let* $g_{ik}(u,v)$ *and* $h_{ik}(u,v)$, *where* $i,k=1,2$, *be defined on a simply-connected* (u,v)*-domain,* R. *Necessary and sufficient in order that* (g_{ik}), (h_{ik}) *be the matrices of the first and second fundamental form of a surface* $X(u,v)$ *of class* C^2 *on* R *is that* (g_{ik}) *be symmetric, positive-definite, of class* C^1, *that* (h_{ik}) *be symmetric and continuous, and that* (g_{ik}), (h_{ik}) *satisfy* (6) *and* (7) *for arbitrary simple closed curves* J *of class* C^1 *in* R.

When these conditions are satisfied, $X(u,v)$ *is uniquely determined up to translations and rotations of the* (x,y,z)*-space.*

This theorem will be deduced from a modification of the standard existence theorem on systems of total differential equations. In fact, this modified theorem on systems of total differential equations leads to (I) in the same way as the standard theorem leads to Bonnet's theorem.

3. Objections, similar to those mentioned above, can be raised against the standard existence theorem on systems of total differential equations. In this theorem, it is assumed that the given data are of class C^1 and, under this hypothesis, the existence of solutions of class C^2 is proved; cf. [8], §1. It would seem more natural to consider only continuous coefficient functions, and to ask whether or not solutions of class C^1 exist. The answer to this question is given by (II) below when the system is linear. In the non-linear case, treated by (III) below, it is difficult to state integrability conditions without making some differentiability assumptions.

In what follows, the tensor sum convention, used above, will continue to be applied to Greek letters, but not to Latin letters. For example, $a_{i\alpha}z^{\alpha}$ is the sum $\sum_{\alpha} a_{i\alpha}z^{\alpha}$, while $a_{ij}z^{j}$ represents one term (not a sum).

The theorem on total differential equations needed in the proof of (I) is as follows:

(II) *Let* $a^i{}_{jk}(u^1,\cdots,u^n)$ *and* $b^i{}_k(u^1,\cdots,u^n)$, *where* $i,j=1,\cdots,m$ *and* $k=1,\cdots,n$, *be continuous functions on the* n*-dimensional cube* R:

$| u^k | \leqq r$, where $k = 1, \cdots, n$. Necessary and sufficient in order that, for every set of constants $c^1, \cdots, c^m$, the system of total differential equations

$$(8) \qquad dz^i = (a^i_{\alpha\gamma} z^\alpha + b^i_\gamma) du^\gamma \qquad (i = 1, \cdots, m)$$

possess a solution $z^1(u^1, \cdots, u^n), \cdots, z^m(u^1, \cdots, u^n)$ of class C^1 on R satisfying the initial conditions

$$(9) \qquad z^i(0, \cdots, 0) = c^i \text{ for } i = 1, \cdots, m,$$

is that the integrability conditions

$$(10) \qquad \int_J a^i_{jk} du^k + a^i_{jl} du^l = \int\!\!\int_S (a^i_{\alpha k} a^\alpha_{jl} - a^i_{\alpha l} a^\alpha_{jk}) du^k du^l,$$

and

$$(11) \qquad \int_J b^i_k du^k + b^i_l du^l = \int\!\!\int_S (a^i_{\alpha k} b^\alpha_l - a^i_{\alpha l} b^\alpha_k) du^k du^l$$

hold whenever J is a simple closed curve of class C^1 in R bounding a domain S in a plane parallel to the (u^k, u^l)-plane. When the integrability conditions (10) *and* (11) *are satisfied, the solution of* (8) *and* (9) *is unique.*

In the proof of (II) and in the deduction of (I) from (II), the following lemma will be useful.

LEMMA. *Let $a_1(u, v)$, $a_2(u, v)$, $c(u, v)$ be continuous functions defined on a simply-connected (u, v)-domain R and let*

$$(12) \qquad \int_J a_\alpha du^\alpha = \int\!\!\int_S c \, du \, dv, \qquad (u^1, u^2) = (u, v),$$

whenever J is a simple closed curve of class C^1 and S is its interior. Then, if $z(u, v)$ is a function of class C^1 on R,

$$(13) \qquad \int_J z a_\alpha du^\alpha = \int\!\!\int_S (a_2 z_u - a_1 z_v + cz) \, du \, dv.$$

4. The derivation formulae (2) and (3) form a system of the type (8), where $m = 9$, $n = 2$ (and $b^i_k \equiv 0$). The proof that (6) and (7) are necessary and sufficient for the corresponding set of integrability conditions (10) can be verified as in the usual proof of the fact that, under standard conditions, (4) and (5) are equivalent to the complete set of integrability conditions (cf., e. g., [1], pp. 175-178). In the case at hand, the legitimacy of the corresponding operations (of considering linear combinations of integrability conditions) follows from the Lemma.

The verification in question can be sketched as follows: The system (2), (3) splits into three identical systems, corresponding to the three components of the vectors X_1, X_2, N. Hence, it is sufficient to consider the systems (2), (3) in which X_1, X_2, N are treated as scalars, z^1, z^2, z^3. The integrability conditions (10) to be considered are therefore satisfied. Two, corresponding to $i = 1, 2$ and $j = 3$, are identical with (6). Two, corresponding to $i = 3$ and $j = 1, 2$, involve

$$\int_J g^{j\alpha} h_{\alpha\beta} du^\alpha, \qquad j = 1, 2,$$

and are seen to be equivalent to (6), by virtue of the Lemma, since det $g_{ik} \neq 0$. The remaining four are of the form

$$\int_J \Gamma^j{}_{ik} du^\alpha = \int\int_S (\Gamma^\alpha{}_{i1}\Gamma^j{}_{\alpha 2} - \Gamma^\alpha{}_{i2}\Gamma^j{}_{\alpha 1} + g^{j\alpha}h_{\alpha 1}h_{i2} - g^{j\alpha}h_{\alpha 2}h_{i1})\, du\, dv,$$

where $i, j = 1, 2$. By virtue of the Lemma, these equations are equivalent to a set involving

$$\int_J g^{-1} g_{\alpha j} \Gamma^\alpha{}_{i\beta} du^\beta, \qquad i, j = 1, 2.$$

It can be verified that two in the last set of equations are trivially satisfied, and that the other two are identical with (7). This proves that the integrability conditions (10), (11) for the system (2), (3) are equivalent to (6) and (7).

The proof of (I) will be omitted, since it follows the usual proof of Bonnet's theorem, except that (II) replaces the standard theorem on systems of total differential equations.

5. The proof of the Lemma will now be given. Let (u_0, v_0) be an (interior) point of R; let $\delta > 0$ be so small that the square

$$S = S_\delta\colon \; |u - u_0| \leqq \delta, \; |v - v_0| \leqq \delta$$

is in R; finally, let $J = J_\delta$ be the boundary of S. In order to prove the Lemma, it is sufficient to show that if the line integral on the left of (13) is divided by δ^2, then, as $\delta \to 0$, the quotient tends to the integrand, at the point (u_0, v_0), of the double integral on the right of (13), and that this limit process is uniform with respect to (u_0, v_0) on any closed set interior to R.

Since $a_1(u, v)$, $a_2(u, v)$ are continuous and $z(u, v)$ is of class C^1, it follows that, for any point (u, v) of S,

$$a_i(u, v) = a_i^0 + o(1). \tag{14}$$

and

$$z(u, v) = z^0 + (w - u_0) z_u{}^0 + (v - v_0) z_v{}^0 + o(\delta), \tag{15}$$

where the superscript 0 on a function symbol signifies that the argument of the function is (u_0, v_0). The o-terms in (14) and (15) refer to $\delta \to 0$ and are uniform with respect to (u_0, v_0) on any closed set interior to R. Hence the line integral in (13) is the sum

$$z^0 \int_J a_\alpha du^\alpha + \int_J \{(u - u_0) z_u{}^0 + (v - v_0) z_v{}^0\}(a_\alpha{}^0 du^\alpha) + o(\delta^2).$$

By (12), the first term in this sum is, uniformly, $z^0(c^0 + o(1))\delta^2$, since $c(u, v)$ is continuous. The second term in the last formula line is $(a_2{}^0 z_u{}^0 - a_1{}^0 z_v{}^0)\delta^2$. These relations imply the assertion to be proved, and complete the proof of the Lemma.

6. Theorem (II) will now be proved. The sufficiency of (10) and (11) for the existence of a solution $z^1(u^1, \cdots, u^n), \cdots, z^m(w^1, \cdots, u^n)$ of class C^1 of the system (8) and (9) on R will be verified first. To this end, successive approximations will be employed.

On R, put

$$z^i{}_0(u^1, \cdots, u^n) = c^i \text{ for } i = 1, \cdots, m. \tag{16}$$

Suppose that the functions $z^1{}_r, \cdots, z^m{}_r$ have been defined and are of class C^1 on R. Put

$$z^i{}_{r+1}(u^1, \cdots, u^n) = \int_{(0,\ldots,0)}^{(u^1,\ldots,u^n)} (a^i{}_{\alpha\gamma} z^\alpha{}_r + b^i{}_\gamma) du^\gamma + c^i \tag{17}$$

for $i = 1, \cdots, m$, where the integral is a line integral along a path consisting of n line segments parallel to the u^1-, u^2-, $\cdots$, u^n-axis, in this order (some of the line segments can reduce to points).

It will be shown that (17) is of class C^1 in R, and that its partial derivatives are given by

$$\begin{aligned} \partial z^i{}_{r+1}/\partial u^j &= a^i{}_{\alpha j} z^\alpha{}_r + b^i{}_j \\ &+ \sum_{k=j+1}^{n} \int_0^{u_k} \{a^i{}_{\alpha k}(\partial z^\alpha{}_r/\partial u^j - a^\alpha{}_{\beta j} z^\beta{}_r - b^\alpha{}_j) \\ &- a^i{}_{\alpha j}(\partial z^\alpha{}_r/\partial u^k - a^\alpha{}_{\beta k} z^\beta{}_r - b^\alpha{}_k)\} dx, \end{aligned} \tag{18}$$

where the argument of the integrand is $(u^1, u^2, \cdots, u^{k-1}, x, 0, \cdots, 0)$. Let $P = (p^1, \cdots, p^n)$ be a point of R and let $\Delta \neq 0$ be such that

$$\Delta z^i{}_{r+1} = z^i{}_{r+1}(p^1, p^{j-1}, p^j + \Delta, p^{j+1}, \cdots, p^n) - z^i{}_{r+1}(p^1, \cdots, p^n) \tag{19}$$

is defined. If, in the expression

$$(20) \qquad \Delta z^i{}_{r+1} + \int\limits_{p^j+\Delta}^{p^j} (a^i{}_{\alpha j} z^\alpha + b^i{}_j)\, dx,$$

the argument of the integrand is $(p^1, \cdots, p^{j-1}, x, p^{j+1}, \cdots, p^n)$, it is clear from (17) that the expression (20) is equal to

$$(21) \qquad \sum_{k=j+1}^{n} \int\limits_{J_k} z^\alpha{}_r (a^i{}_{\alpha j} du^j + a^i{}_{\alpha k} du^k) + \int\limits_{J_k} b^i{}_j du^j + b^i{}_k du^k,$$

where J_k is the boundary of the rectangle in which $u^l = p^l$ if $l < j$ or $j < l < k$, and $u^l = 0$ if $l > k$, while u^j varies from p^j to $p^j + \Delta$ and u^k varies from 0 to p^k. Since it is assumed that $z = z^i{}_r$ is of class C^1 in R for $i = 1, \cdots, m$, the Lemma is applicable to each of the integrals in the first term of (21). Hence, the function (21) is identical with

$$(22) \qquad \sum_{k=j+1}^{n} \int\limits_{p^j}^{p^j+\Delta} \int\limits_{0}^{p^k} \{(a^i{}_{\alpha k} \partial z^\alpha{}_r / \partial u^j - a^i{}_{\alpha j} \partial z^\alpha{}_r / \partial u^k) + z^\alpha{}_r (a^i{}_{\beta j} a^\beta{}_{\alpha k}$$
$$- a^i{}_{\beta k} a^\beta{}_{\alpha j}) + a^i{}_{\alpha j} b^\alpha{}_k - a^i{}_{\alpha k} b^\alpha{}_j\} du^j du^k,$$

where the argument of the integrand is $(p^1, \cdots, p^{j-1}, u^j, p^{j+1}, \cdots, p^{k-1}, u^k, 0, \cdots, 0)$. If the ratio $\Delta z^i{}_{r+1}/\Delta$ is calculated by equating expressions (20) and (22), and Δ then tends to 0, it is seen that $\partial z^i{}_{r+1}/\partial u^j$ exists at P, and that (18) holds if $(p^1, \cdots, p^n)$ is replaced by $(u^1, \cdots, u^n)$. Clearly, (18) implies the continuity of the function $\partial z^i{}_{r+1}/\partial u^j$ on R. Consequently, $z^1{}_{r+1}, \cdots, z^m{}_{r+1}$ are of class C^1 on R.

That the functions (17), where $i = 1, 2, \cdots, m$, converge uniformly, as $r \to \infty$, and that the limits, say $z^1, \cdots, z^m$, are of class C^1 on R and satisfy (8) and (9), follows by standard procedures. Furthermore, these procedures show that the solution of (8) and (9) is unique. The necessary calculations follow, without change, from those of [7], pp. 44-45 and pp. 46-49.

7. It remains to prove that the integrability conditions (10), (11) are satisfied if the existence of a solution $z^1, \cdots, z^m$ of (8) and (9), where the constants $c^1, \cdots, c^m$ are arbitrary, is assumed. This solution is necessarily unique. In fact, the successive approximations (17) are defined and continuous on R and converge uniformly on R (whether or not (10) and (11) are assumed). It follows by standard arguments, involving the use of successive approximations, that if a solution of (8) and (9) exists, it must be the limit of those successive approximations.

The difference of two solutions of (8) is a solution of the homogeneous system obtained by placing $b^i{}_k \equiv 0$ in (8). Hence, the assumption that (8) and (9) have a solution implies that the homogeneous system and (9) have a solution. As above, this solution is necessarily unique. Consequently, if $z^1, \cdots, z^m$ is a solution of the homogeneous system and $z^1, \cdots, z^m$ vanish at some point of R, then they vanish identically. It follows that there exist m solutions of the homogeneous system, and that every solution is a linear combination (with constant coefficients) of these m solutions. Furthermore, if $c^1, \cdots, c^m$ are arbitrary constants and $(u^1{}_0, \cdots, u^n{}_0)$ is an arbitrary point of R, then there exists a solution $z^1, \cdots, z^m$ of the homogeneous system satisfying

$$(23) \qquad z^1(u^1{}_0, \cdots, u^n{}_0) = c^1, \cdots, z^m(u^1{}_0, \cdots, u^n{}_0) = c^m.$$

In order to prove (10), let $(u^1{}_0, \cdots, u^n{}_0)$ be an interior point of R, and let $\delta > 0$ be so small that $J = J_\delta$, the boundary of the square

$$S_\delta: \qquad u^j = u^j{}_0 \text{ if } j \neq k, l; \ | u^k - u^k{}_0 | < \delta \text{ and } | u^l - u^l{}_0 | < \delta,$$

is in R. It is sufficient to show that if the line integral on the left of (10) is divided by δ^2, then the quotient tends, as $\delta \to 0$, uniformly to the integrand, at $(u^1{}_0, \cdots, u^n{}_0)$, of the double integral on the right of (10), and that this limit process is uniform with respect to $(u^1{}_0, \cdots, u^n{}_0)$ on any closed set interior to R.

To this end, let $z^1, \cdots, z^n$ be the solution of class C^1 of (23) and of the homogeneous system obtained from (8) by placing $b^i{}_k \equiv 0$. Then

$$(24) \qquad \int_J z^\alpha(a^i{}_{\alpha k} du^k + a^i{}_{\alpha l} du^l) = 0,$$

and, on S,

$$z^\alpha = c^\alpha + (u^\gamma - u^\gamma{}_0) a^\alpha{}_{\beta\gamma 0} c^\beta + o(\delta) \quad \text{and} \quad a^i{}_{jk} = a^i{}_{jk0} + o(1),$$

where $a^i{}_{jk0} = a^i{}_{jk}(u^1{}_0, \cdots, u^n{}_0)$. The o-terms refer to $\delta \to 0$ and are uniform with respect to $(u^1{}_0, \cdots, u^n{}_0)$ on any closed set interior to R. It follows from (24) that

$$c^\alpha \int_J a^i{}_{\alpha k} du^k + a^i{}_{\alpha l} du^l = c^\beta \int_J (u^\gamma - u^\gamma{}_0) a^\alpha{}_{\beta\gamma 0}(a^i{}_{\alpha k0} du^k + a^i{}_{\alpha l0} du^l) + o(\delta^2).$$

But the right-hand side of this equation is

$$c^\beta (a^i{}_{\alpha l0} a^\alpha{}_{\beta k0} - a^i{}_{\alpha k0} a^\alpha{}_{\beta l0}) \delta^2 + o(\delta^2).$$

If c^α is chosen to be 1 or 0 according as $\alpha = j$ or $\alpha \neq j$, the desired limit relation follows. This proves the necessity of condition (10).

The proof of the necessity of (11) is similar and will be omitted.

8. Theorem (II) has an analogue for non-linear systems of total differential equations.

(III) *Let* $a^{i}{}_{k}(u^1, \cdots, u^n; z^1, \cdots, z^m)$ *be continuous on the parallelepiped* $R: (|z^i| \leqq b,\ i = 1, \cdots, m;\ |u^k| \leqq a,\ k = 1, \cdots, n)$ *and let* $|a^{i}{}_{k}| \leqq M$ *on* R. *Suppose further that the functions* $a^{i}{}_{k}$ *possess continuous partial derivatives with respect to the variable* $z^1, \cdots, z^m$ *on* R, *and that*

$$(25) \qquad \int_J a^{i}{}_{k} du^k + a^{i}{}_{l} du^l = \iint_S (a^{\alpha}{}_{l} \partial a^{i}{}_{k} / \partial z^{\alpha} - a^{\alpha}{}_{k} \partial a^{i}{}_{l} / \partial z^{\alpha}) \, du^k du^l$$

holds whenever J *is a simple closed curve of class* C^1 *in* R *bounding a domain* S *in a plane parallel to the* (u^k, u^l)*-plane. Then, on the parallelepiped* $|u^k| \leqq \min(a, b/nM)$, *the system of total differential equations* $dz^i = a^{i}{}_{\alpha} du^{\alpha}$ *has a unique solution* $z^1(u^1, \cdots, u^n), \cdots, z^m(u^1, \cdots, u^n)$ *of class* C^1, *satisfying the initial conditions* $z^i(0, \cdots, 0) = 0$.

(III) differs from the standard theorems in that in (III) it is not supposed that the functions $a^{i}{}_{k}$ possess partial derivatives with respect to the variables $u^1, \cdots, u^n$.

The proof of (III) will not be given, since it is similar to the proof of (II). It is only necessary to remark that the above Lemma can be replaced by the following fact: The integrability conditions (25) imply that if $z^1(u^1, \cdots, u^n), \cdots, z^m(u^1, \cdots, u^n)$ are functions of class C^1 on a (simply connected) sub-domain of R, then

$$\int_J a^{i}{}_{k} du^k + a^{i}{}_{l} du^l$$
$$= \iint_S \{ (a^{\alpha}{}_{l} - \partial z^{\alpha}/\partial u^l) \partial a^{i}{}_{k}/\partial z^{\alpha} - (a^{\alpha}{}_{k} - \partial z^{\alpha}/\partial u^k) \partial a^{i}{}_{l}/\partial z^{\alpha} \} \, du^k du^l,$$

where the argument of the integrands is $(u^1, \cdots, u^n; z^1(u^1, \cdots, u^n), \cdots, z^m(u^1, \cdots, u^n))$, and J, S have the same significance as above. (The relations (25) represent the particular case in which $z^1(u^1, \cdots, u^n), \cdots, z^m(u^1, \cdots, u^n)$ are constants.)

Part 2. On Torses.

9. Let K denote Gaussian curvature. It is well-known that, under suitable smoothness restrictions, the solutions of the fundamental equations of a surface, when $K \equiv 0$, can be characterized by either of two properties: (i) the surface is isometric with a plane; (ii) the surface is a torse. But the usual presentations either assume more than what is needed to prove the equivalence of the three conditions (i), (ii), $K \equiv 0$ or fail to state the required restrictions. That some smoothness conditions are needed before these three properties become equivalent was shown, by a discussion of part of the problem, by Lebesgue [6], pp. 319-342.

Although the notions of a Gaussian curvature, a torse, and isometric equivalence, can be defined for surfaces of class C^1, this section will be restricted to the consideration of surfaces of class C^2. Before proceeding further, the pertinent definitions for surfaces $X(u, v)$ of class C^2 will be stated:

(α) The Gaussian curvature K is defined as above, in terms of areas of normal maps, and is therefore the quotient $\det h_{ik}/\det g_{ik}$. Accordingly, $K \equiv 0$ is equivalent to

$$h_{11}h_{22} - h_{12}^2 \equiv 0. \tag{26}$$

(β) A surface is called a torse if through every point of it there passes at least one line segment lying on the surface and having the property that the tangent plane does not vary along the segment.

(γ) A surface is said to be isometric with a plane at a point (u_0, v_0) if, in some vicinity of (u_0, v_0), there exists a transformation $u^* = u^*(u, v)$, $v^* = v^*(u, v)$ with the properties that it is of class C^1, that its Jacobian does not vanish, and that the element of arc-length on the surface in the (u^*, v^*)-coordinates is Euclidean, $ds^2 = du^{*2} + dv^{*2}$.

It will be shown that, barring an exceptional case, the properties of being a torse, of being isometric with a plane, and having a curvature $K \equiv 0$, are equivalent for surfaces $X(u, v)$ of class C^2.

The first theorem to be proved is

(IV) *Let $X(u, v)$ be a surface of class C^2. The Gaussian curvature is identically* 0 *if and only if $X(u, v)$ is isometric with a plane at every point.*

The statement to be proved concerning the equivalence of $K \equiv 0$ and of the property of being a torse will involve a qualifying condition:

(V) *Let $X(u, v)$ be a surface of class C^2. The Gaussian curvature is identically* 0 *if $X(u, v)$ is a torse; conversely, if the Gaussian curvature is identically* 0 *and if either every point of $X(u, v)$ is a flat point or every point of $X(u, v)$ is a non-flat point, then $X(u, v)$ is a torse.*

By a flat point (u_0, v_0) is meant a point at which the matrix of the second fundamental form is the zero matrix, that is,

$$h_{11} = h_{12} = h_{22} = 0. \tag{27}$$

It will remain an open question whether or not the restriction in the second part of (V), namely, that every point of $X(u, v)$ be of the same type, flat or non-flat, can be dropped.

10. The part of (IV) which states that $K \equiv 0$ implies that $X(u, v)$ is isometric with a plane at every point will be proved first. A standard proof of this implication involves the introduction of geodesic parallel coordinates in a vicinity of a given point, and an application of (5) in those coordinates. It turns out that this method can be adapted to the more general case of the present assumptions. This adaptation involves the replacement of (5) by its integrated form, (7).

By modifying the procedure applied in [3], it can be shown that, in a vicinity of any (u_0, v_0), new parameters u^*, v^* can be introduced as follows: The transformation $u^* = u^*(u, v)$, $v^* = v^*(u, v)$ is of class C^1 and of non-vanishing Jacobian, and the element of arc-length in terms of u^*, v^* is of the form $ds^2 = du^{*2} + G(u^*, v^*)dv^{*2}$, where G (> 0) is continuous and possesses a continuous second partial derivative with respect to u^*. The function G satisfies the Jacobi equation $\partial^2 G^{\frac{1}{2}}/\partial u^{*2} + KG^{\frac{1}{2}} = 0$ and the initial conditions $G(0, v^*) \equiv 1$, $G_{u^*}(0, v^*) \equiv 0$.

Since $K \equiv 0$, these initial conditions imply that $G(u^*, v^*) \equiv 1$. This proves that $X(u, v)$ is isometric with a plane at any point (u_0, v_0).

11. The remaining part of (IV), which states that $K \equiv 0$ if $X(u, v)$ is isometric with a plane at every point, will now be proved. This statement becomes trivial if the surface has a parametrization (near every point) in which the element of arc-length is Euclidean and (5) is valid in that parametrization. But definition (γ) assumes that the parametrization of the surface in which the element of arc-length is Euclidean is only of class C^1; so that not even the integrated form, (7), of (5) is applicable in this coordinate system. This difficulty will be overcome by using a form of (7) which is independent of the coordinate system, namely, the formula of Gauss-Bonnet.

Let (u_0, v_0) be an arbitrary point of the surface $X(u, v)$ of class C^2. Let u^*, v^* denote the parameters described in the definition (γ), whose existence is assumed. Although the surface in the parametric representation (u^*, v^*) might only be of class C^1, it is clear that the (surface) image of any line segment in the (u^*, v^*)-domain is a geodesic (hence, an arc of class C^2), and conversely.

On surfaces of class C^2, the theorem of Gauss-Bonnet is valid ([9], pp. 43-44; cf. [5], p. 136). Hence, if T is the interior of a geodesic triangle, $\alpha + \beta + \gamma + \int\int_T K d\sigma = 2\pi$, where $d\sigma$ is the element of surface area and

α, β, γ are the exterior angles of the triangle. Consider small geodesic triangles with the point (u_0, v_0) as a vertex. In the (u^*, v^*)-coördinates, such a geodesic triangle is the image of a triangle in the (u^*, v^*)-plane (in particular, the geodesic triangles exist). Since the (u^*, v^*)-coordinates are Euclidean, the angles α, β, γ are also the exterior angles of the triangle in the (u^*, v^*)-plane. Thus $\alpha + \beta + \gamma = 2\pi$, and so $\iint_T K d\sigma = 0$. Since the domain T is arbitrarily small and since K is continuous, it follows that $K(u_0, v_0) = 0$. This completes the proof of (IV).

12. In order to prove the first part of (V), let $X(u, v)$ be a torse of class C^2. Let (u_0, v_0) be an arbitrary point of the surface and let $u = u(t)$, $v = v(t)$ denote a parametric representation of a line segment which lies on the surface, passes through (u_0, v_0) and has the property that the tangent plane does not vary along this segment. Since the surface is of class C^2, and since the vector product $(X_u, X_v) \neq 0$, the functions $u(t)$, $v(t)$ are of class C^2.

If $N(u, v)$ is the unit surface normal, $dN(u(t), v(t))/dt = 0$. It follows therefore by the formula of Rodrigues that $u = u(t)$, $v = v(t)$ is a line of curvature, and that 0 is a principal curvature along the segment. Hence, the Gaussian curvature, being the product of the principal curvatures, vanishes along the segment; in particular, it vanishes at (u_0, v_0). Since (u_0, v_0) is an arbitrary point, the first part of (V) follows.

13. It remains to prove the second part of (V). It can be supposed that the surface is given in the form $z = z(x, y)$, where z is defined in a vicinity of $(x, y) = (0, 0)$ and is of class C^2. The assumption $K \equiv 0$, being equivalent to (26), reduces to

$$(28) \qquad z_{xx} z_{yy} - z_{xy}^2 \equiv 0,$$

since h_{11}, h_{12}, h_{22} are identical with $(1 + z_x^2 + z_y^2)^{-\frac{1}{2}}$ times z_{xx}, z_{xy}, z_{yy}, respectively.

First, if every point is a flat point, that is, if (27) holds for all points (x, y) near $(0, 0)$, then z is a linear function of x and y. Thus the surface is part of a plane, hence a torse.

Next, if $(0, 0)$ is a non-flat point, then at least one of the functions z_{xx}, z_{xy}, z_{yy} does not vanish at $(0, 0)$, and therefore in a vicinity of $(0, 0)$. It can be supposed that $z_{xx}(0, 0) \neq 0$ (otherwise the (x, y)-plane can be subjected to a suitable rotation).

Under the asumption (26), the differential equation of the asymptotic lines, $h_{11}x'^2+2h_{12}x'y'+h_{22}y'^2=0$, reduces to either of the equivalent equations $h_{11}x'+h_{12}y'=0$, $h_{12}x'+h_{22}y'=0$, where the prime denotes differentiation with respect to a parameter which does not occur explicitly. But the last two equations can be written as $z_{xx}x'+z_{xy}y'=0, z_{yx}x'+z_{yy}y'=0$, that is, as $(z_x)'=0$, $(z_y)'=0$; so that $z_x=\text{const.}$, $z_y=\text{Const.}$ Thus the tangent plane does not vary along an asymptotic curve. Conversely, if, along an arc of class C^1, the function z_x (and/or z_y) is constant, then the arc is an asymptotic arc.

Since $z_{xx}\neq 0$ in a vicinity of $(0,0)$, it follows from the theorem on implicit functions that, if t is fixed and $|t|$ is sufficiently small, the equation $z_x(x,y)=z_x(t,0)$ has a solution $x=x(y,t)$ of class C^1 for $|y|$, $|t|$ small. For fixed t, the arc $x=x(y,t)$ is an asymptotic arc, since $z_x(x,y)=\text{const.}$ $(=z_x(t,0))$. The variables t, y can be introduced as new parameters, since $x(0,t)\equiv t$ implies that $\partial(x,y)/\partial(t,y)=1$ at $(t,y)=(0,0)$.

Let $z=z(y;t)=z(x(y,t),y)$. The linear equation

$$a(t)x(y,t)+b(t)y-z(y;t)=c(t), \tag{29}$$

is an identity in y and t if $a(t)=z_x(t,0)$, $b(t)=z_y(t,0)$ and $c(t)=a(t)t-z(0;t)$. (Thus, the surface is the envelope of a 1-parameter sheaf of planes, which is of class C^1.) It follows at once that the asymptotic arc $x=x(y,t)$, for t fixed, is a line segment. In fact, differentiation of (29) with respect to t gives

$$a_t(t)x(y,t)+b_t(t)y=c_t(t),$$

since $ax_t-z_t=0$ is the scalar product of a normal vector and a vector tangent to the (surface) arc $x=x(y,t)$, for y fixed. Since $a_t(t)=z_{xx}(t,0)\neq 0$, the equations

$$a(t)x+b(t)y-z=c(t) \text{ and } a_t(t)x+b_t(t)y=c_t(t)$$

define planes which intersect along the asymptotic arcs $x=x(y,t)$ for t fixed.

This completes the proof of (V). The proof is an adaptation of one given by Hadamard [2], pp. 398-399.

Part 3. The Fundamental Equations of Curves.

14. The fundamental theorem in the differential geometry of curves states that if $\kappa(s)>0$ and $\tau(s)$ are, on some s-interval, functions of class C^1 and C^0, respectively, then there exists, on this interval, a curve $X=X(s)$

of class C^3 for which s, $\kappa(s)$ and $\tau(s)$ become the arc-length, the curvature and the torsion, respectively. Furthermore, this $X = X(s)$ is unique, up to translations and rotations of the (x, y, z)-space.

The proof of this theorem involves the Frenet differential equations,

$$(30) \qquad U_1' = \kappa U_2, \qquad U_2' = -\kappa U_1 + \tau U_3, \qquad U_3' = -\tau U_2,$$

where the prime denotes differentiation with respect to s and U_1, U_2, U_3 are certain unit vectors in the X-space. Since the linear differential equations (30) possess solutions if the coefficient functions κ, τ are merely continuous, there arises the question as to what remains of the fundamental theorem if it is only assumed that κ (> 0) and τ are continuous. In order to answer this question, it will be necessary to extend the definition of torsion.

Let $X = X(s)$ be an arc of class C^2, of non-vanishing curvature, having its arc-length s as parameter. Let $U_1 = X'$, $U_2 = X''/|X''|$ and $U_3 = (U_1, U_2)$, where the symbol $(\ ,\ )$ is that of vector multiplication, be the unit tangent, the principal normal and the binormal unit vectors, respectively. The arc $X(s)$ will be said to possess a torsion $\tau = \tau(s)$ at a point s if

$$(31) \qquad \lim_{\Delta s \to 0} |\Delta\theta/\Delta s|$$

exists, where $\Delta\theta$ is a continuous determination of the angle between the vectors $U_3(s)$ and $U_3(s + \Delta s)$, such that $\Delta\theta = 0$ for $\Delta s = 0$. When the limit (31) exists it will be called the absolute value of the torsion; so that

$$|\tau(s)| = \lim_{\Delta s \to 0} |(U_3(s), U_3(s + \Delta s))/\Delta s|.$$

In order to define the sign of $\tau(s)$, note that the definition, $U_3 = (U_1, U_2)$, of the binormal implies that

$$(32) \qquad (U_3(s), U_3(s + \Delta s)) = [U_1, U_2, \Delta U_2]U_1 - [U_1, U_2, \Delta U_1]U_2,$$

where $\Delta U_i = U_i(s + \Delta s) - U_i(s)$, and $[A, B, C]$ is the scalar triple product of the vectors A, B, C. Since $\Delta U_1 = (\kappa U_2 + o(1))\Delta s$ as $\Delta s \to 0$, the last term of (32) is $o(|\Delta s|)$ as $\Delta s \to 0$. Hence the limit (31) exists if and only if $\lim |[U_1, U_2, \Delta U_2/\Delta s]|$ exists (as $\Delta s \to 0$); in which case the two limits are equal. If the last limit exists and is not 0, the limit also exists if the sign of the absolute value is removed. The sign of $\tau(s)$ can be determined from this limit; thus, if (31) exists, put

$$(33) \qquad \tau(s) = \lim_{\Delta s \to 0} [U_1, U_2, \Delta U_2/\Delta s].$$

It is clear that if the torsion exists, in the above sense, at a point, then its absolute value has the usual geometrical interpretation as the measure of the rate of the twisting of the osculating plane.

Since $U_3 = (U_1, U_2)$ and $U_2 = (U_3, U_1)$, and since U_1 is of class C^1, it is clear that U_2 has a derivative at a given point s if and only if U_3 does. Clearly, a sufficient condition for the existence of the torsion at a point is the existence of U_2' (and/or U_3'); in which case (33) shows that τ is given by the scalar product $U_3 \cdot U_2'$. Furthermore, the Frenet equations (30) hold at such an s. In fact, the first of the equations (30) is the definition of U_2. Since the definition of U_3 implies that U_3' is orthogonal to U_1 and U_3, while $U_3' \cdot U_2 = - U_2' \cdot U_3$, the third of the equations (30) holds. Finally, the second is a consequence of the other two. Conversely, the fundamental theorem can be stated as follows:

(VI) *Let $\kappa(s) > 0$, $\tau(s)$ be continuous functions on an s-interval. Then there exists, on that interval, a curve $X = X(s)$ of class C^2 for which s is arc-length and $\kappa(s)$, $\tau(s)$ are the curvature and the torsion, respectively; in fact, the unit tangent, principal normal, and binormal unit vectors are of class C^1 and satisfy* (30). *The curve $X(s)$ is unique up to translations and rotations of the (x, y, z)-space.*

The proof depends on an adaptation of the proof of the classical theorem. Let $S = (s_{ij})$ denote the three-rowed skew-symmetric matrix in which $s_{12} = \kappa$, $s_{13} = 0$ and $s_{23} = \tau$. If U denotes the matrix whose consecutive columns are U_1, U_2, U_3, then the system of differential equations (30) can be written in the matrix form

$$(34) \qquad U' = - US, \qquad (S^* = - S),$$

when the asterisk denotes transposition. It follows from (31) that $U^{*\prime} = SU^*$, hence $(UU^*)' = - USU^* + USU^* = 0$. Thus, $UU^* = \text{const.}$, where const. denotes a constant matrix, is a first integral of (34).

Since the differential equations (30) are linear, the initial conditions can be assigned arbitrarily, and the corresponding solution will exist on the interval on which the coefficient functions are defined. The same holds at every s if the initial conditions are so chosen that, at an initial $s = s_0$, the vectors U_1, U_2, U_3 form a right-hand orthonormal system; that is, in such a way that the matrix U is an orthogonal matrix of determinant $+1$. Of course, such a solution U is uniquely determined up to multiplication (from the left) by arbitrary constant orthogonal matrices of determinant $+1$, since (30) is a linear system. It is clear that, corresponding to such a solution U_1, U_2, U_3

of (30), the indefinite integral $X(s) = \int U_1 ds + C$ defines an arc satisfying the assertions of (VI). The uniqueness of X (up to rotations and translations) is clear from the proof.

15. *Application to Enneper's theorem.* The advantage of (VI) over the usual formulation of the fundamental theorem of curves is due to the fact that (VI) does not assume that the curve is of class C^3 and contains, nevertheless, the theory of torsion. In this regard, the situation becomes particularly revealing after a perusal of the standard form of Enneper's theorem; cf., e. g., [1], p. 147. In fact, all that can then be stated is that if the Gaussian curvature, K, is non-positive on a surface of class C^4, and if Γ is an asymptotic curve of non-vanishing curvature, then the square of the torsion is identical with $-K$ along Γ. Clearly, the C^4-assumption, which is usually not stated, is unnatural indeed. But it cannot be omitted in the usual treatment, since otherwise the asymptotic curve Γ need not be of class C^3, and so the torsion (in the usual sense) of Γ becomes undefined.

On the other hand, if (VI) is used, the C^4-assumption can be reduced to the assumption that the surface be of class C^3. In fact, if the surface is of class C^3, the asymptotic arc Γ is of class C^2. Since Γ is of non-vanishing curvature, its binormal, U_3, is the surface normal, N, which is of class C^2; hence, (VI) is applicable to Γ.

Despite the fact that, for the existence of the torsion, it is sufficient that U_3 be only of class C^1, the theorem of Enneper becomes meaningless if it is only assumed that the surface is of class C^2. For an asymptotic arc might then be of class C^1; so that the principal normal U_2, hence the binormal U_3, need not be defined. In order to see this, consider the surface $z = g(x) + xy + \frac{1}{2}y^2$, where $g(x)$ is of class C^2. The equations for the asymptotic lines reduce to $g''x'^2 + 2x'y' + y'^2 = 0$, that is, $dy/dx = -1 \pm (1 - g'')^{\frac{1}{2}}$. Clearly, the Gaussian curvature is negative if and only if $1 - g'' > 0$. Let $h(x) = -1 \pm (1 - g'')^{\frac{1}{2}}$; so that

$$g'' = -2h - h^2.$$

Let $h(x)$ be an arbitrary continuous function which vanishes at $x = 0$ and is not of class C^1. Finally, let $g(x)$ be defined by the last formula line (while $g(0)$ and $g'(0)$ are arbitrary). Then the surface, z, is of class C^2 and has a negative curvature, for $|x|$ sufficiently small. It is clear that the asymptotic arcs will not be of class C^2 if $h(x)$ is suitably chosen.

By a suitable choice of an $h(x)$ of class C^1 in the above considerations, it is also seen that, on surfaces of class C^3, the set of those points of on asymptotic arc at which its curvature is 0 can be, for example, a nowhere dense, perfect set.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] A. Duschek and W. Mayer, *Lehrbuch der Differentialgeometrie*, vol. 1, Leipzig and Berlin, 1930.

[2] J. Hadamard, *Cours d'analyse*, vol. 1, Paris, 1927.

[3] P. Hartman, "On the local uniqueness of geodesics," *American Journal of Mathematics*, vol. 72 (1950), pp. 723-730.

[4] ——— and A. Wintner, "On the embedding problem in differential geometry," *ibid.*, vol. 72 (1950), pp. 553-564.

[5] E. R. van Kampen, "The theorems of Gauss-Bonnet and Stokes," *ibid.*, vol. 60 (1938), pp. 129-138.

[6] H. Lebesgue, "Intégrale, longueur, aire," *Annali di Matematica Pura ed Applicata*, ser. 3, vol. 7 (1902), pp. 231-359.

[7] W. Nikiliborc, "Sur les équations linéaires aux différentielles totales," *Studia Mathematica*, vol. 1 (1929), pp. 41-49.

[8] F. Schur, "Ueber den analytischen Charakter der eine endliche continuierliche Transformationsgruppe darstellenden Functionen," *Mathematische Annalen*, vol. 41 (1893), pp. 509-538.

[9] H. Weyl, "Ueber die Bestimmung einer geschlossenen konvexen Fläche durch ihr Linienelement," *Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich*, vol. 61 (1916), pp. 40-72.

LAMÉ COORDINATES IN HILBERT SPACE.*

By Philip Hartman and Aurel Wintner.

1. In an n-dimensional unitary space, let $\xi = (\xi_1, \xi_2, \cdots, \xi_n)$ be a point not on a parameter plane (so that $\xi_k \neq 0$ for $k = 1, \cdots, n$), and let $\alpha_1 < \alpha_2 < \cdots < \alpha_n$. Then the equation

$$(1) \qquad \sum_{k=1}^{n} | \xi_k |^2 / (\alpha_k - \lambda) = 1$$

has n distinct real roots, say, $\lambda_1 < \lambda_2 < \cdots < \lambda_n$, which separate, and are separated by, the numbers $\alpha_1, \cdots, \alpha_n$. This is the classical definition of the Lamé coordinates $\lambda_1, \cdots, \lambda_n$ of ξ. The following considerations will be made possible by the fact that it is possible to introduce the Lamé coordinates by quite another, though equivalent, definition, as follows:

Let (x, y) denote the scalar product, $x_1\bar{y}_1 + \cdots + x_n\bar{y}_n$, of x and y, and let $\phi = c^{-\frac{1}{2}}\xi$, where $c = (\xi, \xi)$ is the squared length of ξ. If H denotes the diagonal matrix having the diagonal elements $\alpha_1, \alpha_2, \cdots, \alpha_n$, and if I is the unit matrix, then (1) can be written in the form $((H - \lambda I)^{-1}\phi, \phi) = -1/c$. Let E denote the matrix of the linear transformation

$$(2) \qquad Ex = (x, \phi)\phi;$$

so that E is a one-dimensional projection. It turns out that the eigenvalues of the Hermitian matrix $H + cE$ are the Lamé coordinates, $\lambda_1, \cdots, \lambda_n$, of ξ.

In this characterization of the Lamé coordinates of ξ, the restriction that all coordinates ξ_k of ξ be different from 0 can be omitted. This generalization merely allows the possibility that some of the Lamé coordinates $\lambda_1, \cdots, \lambda_n$ of ξ and some of the numbers $\alpha_1, \cdots, \alpha_n$ are equal. (Of course, there still remains the difficulty that $(\xi_1, \cdots, \xi_n)$ and $(e^{i\theta_1}\xi_1, \cdots, e^{i\theta_n}\xi_n)$, for arbitrary real $\theta_1, \theta_2, \cdots, \theta_n$, have the same Lamé coordinates.) It is clear that, in this definition, H need not be a diagonal matrix with distinct roots but can be an arbitrary Hermitian matrix. In the latter case, the number α occurs at least $k - 1$ times among the Lamé coordinates of every element ξ, if the characteristic number $\lambda = \alpha$ of H is of multiplicity k.

2. The object of this note is to carry over these ideas to Hilbert's space. Let H denote an arbitrary Hermitian matrix, bounded in Hilbert's space,

---

* Received April 5, 1950.

ϕ a vector of length 1 in this space, and E the corresponding projection (2). Associated with H and ϕ is a bounded Hermitian matrix, say K, which is the projection of H on the hyperspace orthogonal to ϕ. Thus, $K = (I - E)H(I - E)$, where I is the unit matrix. For example, if the coordinates of ϕ are $(1, 0, 0, \cdots)$, then K is obtained by removing from H the first row and first column. The matrix K is considered as operating on the hyperplane orthogonal to ϕ. Clearly,

$$K = (I - E)(H + cE)(I - E), \text{ where } -\infty < c < \infty.$$

It was recently shown ([3], p. 867) that the spectra of K and H separate each other; in the sense that if λ_1 and λ_2, where $\lambda_1 < \lambda_2$, are two points of the spectrum of K (or H), then the interval $\lambda_1 \leqq \lambda \leqq \lambda_2$ contains at least one point of the spectrum of H (or K). The same holds in the limiting case $\lambda_1 = \lambda_2$, if $\lambda = \lambda_1 = \lambda_2$ is in the point spectrum of K (or H) with a multiplicity $m > 1$; in which case λ is in the point spectrum of H (or K) with at least the multiplicity $m - 1$.

These separation theorems imply that the cluster points of the spectra of K and H are identical. This corollary is contained in a theorem of Weyl [9], p. 384, also, since the difference $K - H$ is completely continuous.

The above comments make clear the content of the following theorems:

(I) *Let the spectrum of K contain the points λ_1, λ_2 but no point of the interval $\lambda_1 < \lambda < \lambda_2$. Then, to every λ on this interval, there belongs a unique real $c = c(\lambda)$; corresponding to which λ is in the spectrum of $H + cE$. The function $c = c(\lambda)$, given by*

$$c(\lambda) = -1/((H - \lambda I)^{-1}\phi, \phi), \tag{3}$$

is regular-analytic and increasing for $\lambda_1 < \lambda < \lambda_2$.

That (3) is meaningful, that is, that $((H - \lambda I)^{-1}\phi, \phi) \neq 0$, will be shown in the proof of (I). It is understood that $c(\lambda) = 0$ if λ is in the spectrum of H.

(II) *Let the conditions of* (I) *hold and let $\lambda = \lambda_1$ be an isolated point of the point spectrum of K, hence of $H + cE$. Let m, $m(c)$ denote the multiplicity of λ_1 in the point spectrum of K, $H + cE$, respectively, so that $0 \leqq m - 1 \leqq m(c) \leqq m + 1$. Finally, let $c_1 = \lim c(\lambda)$ as $\lambda \to \lambda_1 + 0$, where $-\infty \leqq c_1 < \infty$. Then $c_1 = -\infty$ holds if and only if $m(c) = m - 1$ is an identity for $-\infty < c < \infty$.*

(II bis) *If $c_1 > -\infty$, then $m(c) = m$ or $m(c) = m + 1$ according as $c \neq c_1$ or $c = c_1$.*

3. Proof of (I). If λ is on the interval (λ_1, λ_2), then there is at most one c corresponding to which λ is in the spectrum of $H + cE$. For suppose, if possible, that there exist two distinct values c_1, c_2 with the property that λ is in the spectra of both $H + c_1E$ and $H + c_2E$. Let x_1, x_2 denote normalized eigenfunctions satisfying

$$(4) \qquad (H + cE)x = \lambda x$$

for $c = c_1, c_2$, respectively. The elements x_1, x_2 are either linearly independent or dependent. In the first case, let x denote a linear combination of x_1, x_2 of length 1 and orthogonal to ϕ, that is, $|x| = 1$ and $Ex = 0$. It then follows from (4) that $Hx = \lambda x$. In the second case, it can be supposed that $x = x_1 = x_2$. Then (4) shows that $(c_1 - c_2)Ex = 0$; hence $Ex = 0$, since $c_1 \neq c_2$. Thus, in either case, there exists an element x, of length 1, satisfying $Hx = \lambda x$ and $Ex = 0$. But the definition of K implies that

$$(5) \qquad Kx = \lambda x,$$

which contradicts the assumption that no point of the interval (λ_1, λ_2) is in the spectrum of K.

The separation theorem, quoted before (I), shows that the interval (λ_1, λ_2) contains at most one point of the spectrum of H. If such a point exists. denote it by λ_0. If $\lambda \neq \lambda_0$, let $x = (H - \lambda I)^{-1}\phi$. Then x exists and is not 0. Furthermore, $(x, \phi) = ((H - \lambda I)^{-1}\phi, \phi) \neq 0$. For otherwise $(I - E)x = x$ holds, hence $(K - \lambda(I - E))x$ becomes $(I - E)(H - \lambda I)x$, which is $(I - E)\phi$ and therefore 0. But this contradicts the assumption that λ is not in the spectrum of K. Consequently, (3) is meaningful if $\lambda = \lambda_0$. The regular-analyticity and monotony of $c(\lambda)$ at $\lambda \neq \lambda_0$ follow from the identity

$$(6) \qquad ((H - \lambda I)^{-1}\phi, \phi) = \int_{-\infty}^{+\infty} (\mu - \lambda)^{-1} d\rho(\mu),$$

where $\rho(\mu)$ is the non-decreasing function on $-\infty < \mu < \infty$ defined by $\rho(\mu) = (E(\mu)\phi, \phi)$ in terms of the resolution of the identity $E(\mu)$ belonging to H.

The definitions of x, E and $K = (H - \lambda I)^{-1}\phi$ show that $(H - \lambda I + cE)x = \phi + cEx = \phi + c((H - \lambda I)^{-1}\phi, \phi)\phi$. If c is given by (3), it is seen that (4) holds; so that λ is in the point spectrum of $H + cE$.

It remains to show that if λ_0 exists, then $c(\lambda)$ is regular-analytic and increasing at $\lambda = \lambda_0$. (Of course, $c(\lambda_0) = 0$.) To this end, let $c_0 \neq 0$ be a real number; so that λ_0 (and hence any λ near λ_0) is not in the spectrum of $H_0 = H + c_0E.$. Since the relation between K and $H + cE$ is independent of c, it follows that if λ is on the interval (λ_1, λ_2), then λ is in the spectrum of $H_0 + cE$ if and only if $c = c(\lambda) - c_0$. Thus, for λ near λ_0,

$$(7) \qquad c(\lambda) - c_0 = -1/((H_0 - \lambda I)^{-1}\phi, \phi).$$

This completes the proof of (I).

4. Proof of (II). It will first be shown that if $m(c) > m - 1$ for one value of c, say for $c = c_0$, then $m(c) > m - 1$ for all values of c.

Let $x_1, \cdots, x_m$ denote m ortho-normalized solutions of (4), where $c = c_0$ and $\lambda = \lambda_1$. If $(x_j, \phi) = 0$ for $j = 1, \cdots, m$, then $x_1, \cdots, x_m$ are solutions of (4), with $\lambda = \lambda_1$, implying that $m(c) \geqq m$ holds for all values of c. It can, therefore, be supposed that $(x_1, \phi) \neq 0$. It can also be supposed that $(x_j, \phi) = 0$ for $j = 2, \cdots, m$. For otherwise x_j, where $j = 2, \cdots, m$, can be replaced by a suitable linear combination of x_j and x_1 (then x_1 becomes replaced by an element orthogonal to $x_2, \cdots, x_m$). Consequently, the elements $x_2, \cdots, x_m$ are also solutions of (5), with $\lambda = \lambda_1$. Since $\lambda = \lambda_1$ has a multiplicity m in the point spectrum of K, there exists a normalized element $x = y$ satisfying (5) with $\lambda = \lambda_1$, $(y, \phi) = 0$ and $(y, x_j) = 0$ for $j = 2, \cdots, m$.

On the other hand, the definition of K and the relations $(y, \phi) = 0$ and (5), with $\lambda = \lambda_1$, imply that there exists a constant k satisfying $(H - \lambda I)y = k\phi$. It can be supposed that $k \neq 0$, for otherwise $y, x_2, \cdots, x_m$ are m ortho-normalized solutions of (4), with $\lambda = \lambda_1$, for every c, and the assertion is proved.

If $k \neq 0$, it is easily verified that $\alpha y + x_1$ is a solution of (4), where $\lambda = \lambda_1$ and $\alpha = (c_0 - c)(x_1, \phi)/k$. Clearly, $\alpha y + x_1$ is orthogonal to $x_2, \cdots, x_m$; and $\alpha y + x_1 \neq 0$, since $(\alpha y + x_1, \phi) = (x_1, \phi) \neq 0$. Hence (4), with $\lambda = \lambda_1$, has at least m linearly independent solutions for every c.

5. The assertions of (II) can now be proved as follows:

First, it will be shown that if $m(c) = m - 1$ holds for all values of c, then $c_1 = -\infty$.

Suppose, if possible, that $c_1 > -\infty$. Let $y, x_2, \cdots, x_m$ denote m ortho-normalized solutions of (5), with $\lambda = \lambda_1$, chosen so as to satisfy $(y, \phi) = (x_j, \phi) = 0$ and $(H - \lambda_1 I)x_j = 0$ for $j = 2, \cdots, m$. Let $x(\lambda)$ denote a normalized solution of (4), for $c = c(\lambda)$ and λ $(> \lambda_1)$ near λ_1. There exists a sequence of λ-values $\lambda^1, \lambda^2, \cdots$ satisfying $\lambda^n > \lambda^{n+1} \to \lambda_1$ as $n \to \infty$, and $x_1 = \lim x(\lambda^n)$ exists in the weak sense. Since, as $n \to \infty$,

$$(8) \qquad (H - \lambda_1 I + c_1 E)x(\lambda^n) = (\lambda^n - \lambda_1)x(\lambda^n) + (c_1 - c(\lambda))x(\lambda^n)$$

tends to 0 in the strong sense, the limit x_1 is not 0. For, if it were, it would follow from Weyl's criterion ([9], p. 378) that $\lambda = \lambda_1$ is a cluster point of the spectrum of $H + c_1E$, and hence of K. Consequently, $x = x_1 \neq 0$ satisfies (4) for $c = c_1$ and $\lambda = \lambda_1$. Also, $(x_1, x_j) = 0$ for $j = 2, \cdot \cdot \cdot, m$. In fact, $(x(\lambda^n), x_j) = 0$ for $j = 2, \cdot \cdot \cdot, m$ and $n = 1, 2, \cdot \cdot \cdot$, since $x(\lambda^n)$ and x_j satisfy (4) for $c = c(\lambda^n)$, but for the distinct λ-values, $\lambda = \lambda^n$ and λ_1, respectively. Since this implies that $m(c_1) \geqq m$, it contradicts the assumption that $m(c) = m - 1$ for every c.

In order to prove the converse, let $c_1 = -\infty$. Suppose, if possible, that $m(c) \geqq m$ for all values of c. The assumption that $\lambda = \lambda_1$ is an isolated point of the spectrum of H shows that the monotone function $\rho(\mu)$ in (6) is constant near $\mu = \lambda_1$, except for a possible jump at $\mu = \lambda_1$.

It follows from (3) that $c_1 = -\infty$ implies that $((H - \lambda I)^{-1}\phi, \phi)$ has a finite limit, viz. 0, as $\lambda \rightarrow \lambda_1 + 0$. Hence, the monotone function $\rho(\mu)$ has no jump at $\mu = \lambda_1$. Thus, ϕ is orthogonal to the linear space spanned by the solutions of $(H - \lambda_1 I)x = 0$. Let $x_1, \cdot \cdot \cdot, x_m$ denote m ortho-normalized solutions of this equation. Then $x_1, \cdot \cdot \cdot, x_m$ are solutions of (5) with $\lambda = \lambda_1$.

Since $\rho(\mu)$ is constant near $\mu = \lambda_1$, it follows that there exists an x, orthogonal to $x_1, \cdot \cdot \cdot, x_m$, satisfying the equation $(H - \lambda_1 I)x = \phi$. In fact, x is given by

$$x = \{\int_{-\infty}^{+\infty} (\mu - \lambda_1)^{-1} dE(\mu)\}\phi \tag{9}$$

(cf., e.g., the Appendix below). Clearly, x satisfies $(x, \phi) = 0$, since $((H - \lambda I)^{-1}\phi, \phi) \rightarrow 0$ as $\lambda \rightarrow \lambda_1 + 0$. Furthermore, it is seen from the definition of K that (5) holds for $\lambda = \lambda_1$. It is also clear that $x_1, \cdot \cdot \cdot, x_m$ and x are linearly independent, since $(H - \lambda_1 I)x_j = 0$ for $j = 1, \cdot \cdot \cdot, m$, while $(H - \lambda_1 I)x = \phi$. But this contradicts the assumption that the multiplicity of $\lambda = \lambda_1$ in the point spectrum of K is only m. Thus the proof of (II) is complete.

6. Proof of (II bis). Let $c_1 > -\infty$; so that $m(c) \geqq m$ for all values c. Let $x_1(c), x_2, \cdot \cdot \cdot, x_m$ denote m ortho-normal solutions of (4), with $\lambda = \lambda_1$, satisfying $(x_j, \phi) = 0$ for $j = 2, \cdot \cdot \cdot, m$. If λ $(> \lambda_1)$ is near λ_1, let $x(\lambda)$ denote a normalized solution of (4), where $c = c(\lambda)$. Then $x(\lambda)$, $x_1(c(\lambda))$, $x_2, \cdot \cdot \cdot, x_2$ form an ortho-normal set. Let $y(\lambda) = \alpha x(\lambda) + \beta x_1(c(\lambda))$, where $\alpha = \alpha(\lambda)$ and $\beta = \beta(\lambda)$, denote a normalized linear combination of $x(\lambda)$ and $x_1(c(\lambda))$ satisfying $(y(\lambda), \phi) = 0$. There exists a sequence of λ-values satisfying $\lambda^n > \lambda^{n+1} \rightarrow \lambda_1$ as $n \rightarrow \infty$, and $y_1 = \lim y(\lambda^n)$ exists in the weak sense. Since $|\alpha|^2 + |\beta|^2 = 1$, it follows that

$$(H-\lambda_1 I)y(\lambda^n) = (H-\lambda_1 I + c(\lambda^n)E)y(\lambda^n) = \alpha(\lambda^n - \lambda_1)x(\lambda^n)$$

tends to 0 in the strong sense, as $n \to \infty$. Consequently, the limit y_1 of $y(\lambda^n)$ is not 0. For, if it were, it would follow from Weyl's criterion used above, that $\lambda = \lambda_1$ is in the essential spectrum of H.

Let $y = y_1/|y_1|$; so that $y, x_2, \cdots, x_m$ form an ortho-normal set, each member of which is orthogonal to ϕ and satisfies (4), with $\lambda = \lambda_1$, for every value of c. By repeating the above argument involving (8), it follows that there exists an element x_1 orthogonal to $y, x_2, \cdots, x_m$ and satisfying (4), with $\lambda = \lambda_1$ and $c = c_1$. Hence, $m(c_1) = m + 1$.

It remains to show that there cannot exist another value of c, say c_2, distinct from c_1 and satisfying $m(c_2) = m + 1$.

Suppose the contrary. Let $x(c_1), y, x_2, \cdots, x_m$ and $x(c_2), y, x_2, \cdots, x_m$ denote ortho-normal sets of solutions of (4), with $\lambda = \lambda_1$, belonging to $c = c_1$ and c_2, respectively. Whether or not $x(c_1)$ and $x(c_2)$ are linearly independent, they have a normalized linear combination, say x^1, which is orthogonal to ϕ; cf. the proof of (I), above. But then $x^1, y, x_2, \cdots, x_m$ are solutions of (5), with $\lambda = \lambda_1$. Since this is a contradiction, the proof of (II bis) is complete.

7. The signatures of the Lamé quadrics in Hilbert space. In the notation of §1, let H be an n-rowed diagonal matrix with $\alpha_1 < \cdots < \alpha_n$ as diagonal elements, and let $\lambda_1 < \cdots < \lambda_n$ be the Lamé coordinates of $\xi = (\xi_1, \cdots, \xi_n)$, where $\xi_1 \neq 0, \cdots, \xi_n \neq 0$. Then, if $k = 1, \cdots, n$, the number of the negative characteristic numbers of $H - \lambda_k I$ is $k - 1$. Correspondingly, for $\lambda = \lambda_1, \cdots, \lambda_n$ the quadric (1) attains each of the n types possible for a (real) central quadric (namely, the $n - 1$ types of hyperboloids and the real ellipsoid). In what follows, this fact concerning consecutive signatures will be extended in the direction of (I)-(II bis) to the case of Hilbert's space.

Let H be a bounded self-adjoint matrix having only simple points (if any) in its point spectrum, and ϕ a vector of length 1 in Hilbert space with the property that no characteristic vector of H is orthogonal to ϕ. Thus, $x = 0$ is implied by $(x, \phi) = 0$ and $Hx = \lambda x$. These assumptions on H and ϕ imply that, for a fixed value of c,

(*) $\qquad H + cE$ does not have a multiple eigenvalue.

For if (*) is false, then there exists a $\lambda = \lambda_c$ corresponding to which (4) has two linearly independent solutions, say x_1 and x_2. Hence, a suitable linear combination x of x_1 and x_2 will be such as to satisfy both $x \neq 0$ and $(x, \phi) = 0$. It follows therefore from (2) that $Ex = 0$, and from (4) that

$Hx = \lambda x$. But the existence of such an x contradicts the assumptions concerning H and ϕ. This proves (*).

Let $c > 0$ and $\xi = c\phi$, so that the "Lamé coordinates of ξ" correspond to the points of the spectrum of $H + cE$. In view of the remarks preceding the statement of (I), the set of the cluster points of the spectrum of $H + cE$, for a fixed c, is identical with the set of cluster points of the spectrum of H. Let λ' and λ'' respectively denote the least and the greatest cluster point of the spectrum of H. Then (I) implies that the (empty, finite or denumerable) set of points $\lambda_1 < \lambda_2 < \cdots$ of the spectrum of $H + cE$ ("Lamé coordinates of ξ") which are less than λ' are determined by the equation $((H - \lambda I)^{-1}\phi, \phi) = -1/c$. Clearly, $-((H - \lambda I)^{-1}\phi, \phi)^{-1} \rightarrow +0$ as $\lambda \rightarrow -\infty$. Also, if λ_0 $(< \lambda')$ is a point of the spectrum of H, then it follows from (I) that $-((H - \lambda I)^{-1}\phi, \phi)^{-1}$ tends to ∞ or $-\infty$ according as $\lambda \rightarrow \lambda_0 - 0$ or $\lambda \rightarrow \lambda_0 + 0$.

These facts, when combined with (*), show that the number of negative points in the spectrum of $H - \lambda_k I$ is $k - 1$. A corresponding statement applies to those "Lamé coordinates of ξ" which exceed λ''. Finally, it is clear from the definition of λ', λ'' that if $\lambda' < \lambda''$, and if λ satisfies $\lambda' < \lambda < \lambda''$ and is in the spectrum of $H + cE$, then there is an infinite number of positive and of negative points in the spectrum $H - \lambda I$.

8. An application to differential operators. Let $p(t)$ be a positive, and $q(t)$ a real-valued, continuous function for $0 \leqq t < \infty$, and let the differential equation

$$(px')' + (\lambda + q)x = 0 \tag{9}$$

be of *Grenzpunkt* type. Then (9) and a homogeneous boundary condition, say

$$x(0)\sin\alpha - p(0)x'(0)\cos\alpha = 0, \tag{10}$$

determine a self-adjoint eigenvalue problem with a real spectrum $S(\alpha)$. Let $T(x, \alpha) = -(px')' - qx$ denote the self-adjoint operator associated with (10); so that the spectrum of $T(x, \alpha)$ is $S(\alpha)$. The set of cluster points S' of $S(\alpha)$ is independent of α; cf. [10], p. 251. Suppose that some real λ-value, say $\lambda = 0$, is not in S'. Then there is a value α_0 of α (unique mod π), corresponding to which $\lambda = 0$ belongs to $S(\alpha_0)$; cf. [1].

Let $x = \phi(t)$ denote a normalized solution of (9) and (10) for $\lambda = 0$ and $\alpha = \alpha_0$; so that $T(\phi, \alpha_0) = 0$. On the sub-domain of $T(x, \alpha_0)$ orthogonal to ϕ, the operator $T(x, \alpha_0)$ possesses a bounded inverse K, since $\lambda = 0$ is an isolated point of $S(\alpha_0)$. If $\alpha \not\equiv \alpha_0 \pmod{\pi}$, then $T(x, \alpha)$ has a bounded inverse $H(\alpha)$, which is an integral operator; cf. [2].

Let E be the projection of the Hilbert space $L^2(0,\infty)$ on the 1-dimensional linear subspace containing ϕ and let $H = H(\alpha_0 + \frac{1}{2}\pi)$. It is easily verified, from the form of the Green kernel belonging to $H(\alpha)$, that $H(\alpha) = H + cE$, where $c = c(\alpha)$ is a constant (independent of α) multiple of $\cot(\alpha_0 - \alpha)$ for $\alpha \not\equiv \alpha_0 \pmod{\pi}$. Furthermore, $K = (I - E)H(\alpha)(I - E)$.

These facts, when combined with theorems (I) and (II), imply the results of [4] concerning the separation of the spectral sets $S(\alpha)$, $S(\beta)$ for $\alpha \not\equiv \beta \pmod{\pi}$.

9. Another illustration: Mathieu functions. For a fixed $q > 0$, let $\lambda_0(q) < \lambda_1(q) < \cdots$ denote the λ-values corresponding to which Mathieu's differential equation,

$$x'' + (\lambda + 16q\cos^2 t)x = 0, \tag{11}$$

has a solution ($\not\equiv 0$) of period π or 2π. It will be shown that the results of Sections 2 and 7 imply that

(i) if $q > 0$ and λ are fixed, then (11) cannot possess two linearly independent solutions of period π or 2π; and that

(ii) according as n in $\lambda = \lambda_n(q)$ is of the form $n = 4k$, $4k + 3$, $4k + 2$, $4k + 1$, the differential equation (11) possesses a solution of the form

$$(12_1) \quad x(t) = \tfrac{1}{2}x + \sum_{m=1}^{\infty} x_m \cos 2mt; \qquad (12_2) \quad x(t) = \sum_{m=1}^{\infty} x_m \sin 2mt,$$

$$(13_1) \quad x(t) = \sum_{m=0}^{\infty} x_m \cos(2m+1)t; \qquad (13_2) \quad x(t) = \sum_{m=0}^{\infty} {}_m \sin(2m+1)t,$$

respectively.

(i) is a theorem of Ince [5]; while (ii) can be derived as a consequence of an asymptotic formula obtained by Ince [6] for $\lambda_k(q)$ as $q \to \infty$ (when k is fixed). Actually, it is well known that if $\lambda = \lambda_0(q)$, then (11) has a solution of the form (12_1); that if λ is $\lambda_{4n+1}(q)$ or $\lambda_{4n+2}(q)$, then (11) has a solution of period 2π (but not of period π), so that the corresponding periodic solution is of one of the forms (13_1) or (13_2); that if λ is $\lambda_{4n+3}(q)$ or $\lambda_{4n}(q)$, then the periodic solution of (11) is of one of the forms (12_1) or (12_2). The point of the assertion (ii) is that $\lambda = \lambda_{4n+1}(q)$ corresponds to a periodic solution of the form (13_1), and $\lambda = \lambda_{4n+2}(q)$ to one of the form (13_2), etc.

In the following proof of (i) and (ii), it will be necessary to use the results of Sections 2 and 7 for the case of a self-adjoint matrix H which is not bounded. The nature of the above proofs makes it clear that this extension is obvious.

Since the factor $\cos 2x$ in (11) is even, any real periodic solution can be assumed to be either even or odd. A comparison of coefficients shows that if (11) has a periodic solution, it can be supposed to have one of the forms (12_1), (12_2), (13_1), (13_2).

Let $H = (h_{nm})$, where $m, n = 0, 1, \cdots$, be the Jacobi matrix in which h_{nm} is $4n^2$, $-8q$ or 0 according as $n = m$, $n - m = \pm 1$ or $|n - m| > 1$. Let ϕ be the vector having the coordinates $(1, 0, 0, \cdots)$ in Hilbert space. Then (2) shows that E is the matrix (e_{nm}) in which every e_{nm}, except $e_{00} = 1$, is 0. Correspondingly, K is the matrix which results if the first row and first column of H are removed; that is, $K = (h_{mn})$, where $m, n = 1, 2, \cdots$. A comparison of coefficients shows that (11) has a solution of the form (12_1) if and only if there exists a point $x = (x_0, x_1, \cdots) \neq 0$ of Hilbert space satisfying

$$(H - \lambda I + \tfrac{1}{2}\lambda E)x = 0; \tag{14}$$

and that (11) has a solution of the form (12_2) if and only if there exists a point $x = (x_1, x_2, \cdots) \neq 0$ in the Hilbert hyperplane $x_0 = 0$ satisfying

$$(K - \lambda(I - E))x = 0. \tag{15}$$

The considerations of Section 7 (cf. (*)) show that $H + cE$ and K cannot have a common eigenvalue if no eigenvector $x = (x_0, x_1, \cdots)$ of H is orthogonal to ϕ. But the form of the equations $(H - \lambda I)x = 0$ shows that if $(x, \phi) = 0$, that is, if $x_0 = 0$, then $x_1 = x_2 = \cdots = 0$. Hence, there exists no value of λ corresponding to which both (14) and (15) have non-trivial solutions

It follows from (I) and (II)-(II bis), Section 2, that the values of λ corresponding to which (14) has non-trivial solutions, separate and are separated by those for which (15) has non-trivial solutions. Hence, the periodic solutions of (11) are of the form (12_1) or (12_2) according as n in $\lambda = \lambda_n(q)$ is of the form $n = 4k$ or $n = 4k + 3$.

In order to consider solutions of the form (13_1) or (13_2), let $H = (h_{mn})$, where $m, n = 0, 1, \cdots$, be the Jacobi matrix defined as follows: If $n \neq 0$ and $m \neq 0$, then h_{mn} is $(2n + 1)^2$, $-8q$ or 0 according as $n = m$, $n = m \pm 1$ or $|n - m| > 1$; while if either $n = 0$ or $m = 0$, then h_{mn} is $1 - 8q$, $-8q$ or 0 according as $n = m$, $n = m \pm 1$ or $|n - m| > 1$. As before, let $\phi = (1, 0, 0, \cdots)$. A comparison of coefficients shows that (11) has a solution of the form (13_1) if and only if there exists a point $x = (x_0, x_1, x_2, \cdots) \neq 0$ in Hilbert space satisfying

$$(H - \lambda I)x = 0; \tag{16}$$

and that (11) has a solution of the form (13_2) if and only if there exists in the same space a non-trivial solution x of

$$(H - \lambda I + 16qE)x = 0. \tag{17}$$

As above, (16) and (17) cannot have non-trivial solutions for the same value of λ. It follows that the eigenvalues of H and $H + 16qE$ separate each other; and that, since $c = 16q > 0$, the least eigenvalue of $H + 16qE$ is greater than that of H. Hence, the periodic solutions of (11) are of the form (13_1) or (13_2) according as n in $\lambda = \lambda_n(q)$ is of the form $n = 4k + 1$ or $n = 4k + 2$.

Appendix.

Let A be a completely continuous matrix; $\lambda_1, \lambda_2, \cdots$ its eigenvalues; $\phi_1, \phi_2, \cdots$ a corresponding set of eigenfunctions; y a point of Hilbert's space; finally,

$$y \sim \sum_{k=1}^{\infty} a_k \phi_k \tag{1}$$

the "Fourier series" of y. Picard [7] has shown that the infinite system of linear equations

$$Ax = y \tag{2}$$

has a solution (in Hilbert space) if and only if

$$\sum_{k=1}^{\infty} | a_k |^2 / | \lambda_k |^2 < \infty; \tag{3}$$

in which case

$$x \sim \sum_{k=1}^{\infty} a_k \phi_k / \lambda_k \tag{4}$$

is a solution of (2), the (unique) solution of least length. In (3), it is understood that if $\lambda_k = 0$ for some k, then $| a_k |^2 / | \lambda_k |^2$ denotes ∞ or 0 according as $a_k \neq 0$ or $a_k = 0$.

The object of this Appendix is to point out the corresponding criterion for the solvability of (2) when A is not completely continuous, but just bounded, say.

Suppose first that A is self-adjoint and let $E(\lambda)$, where $-\infty < \lambda < \infty$, be its resolution of the identity. The equations (2) are solvable if and only if

$$\int_{-\infty}^{+\infty} \lambda^{-2} d \mid E(\lambda) y \mid^2 < \infty; \tag{5}$$

in which case

$$x = \{ \int_{-\infty}^{+\infty} \lambda^{-1} dE(\lambda) \} y \tag{6}$$

is a solution of (2), the solution of least length. This follows from the fact that a necessary condition for the solvability of (2) is that y be on the manifold M orthogonal to the solutions of the homogeneous equations $Ax = 0$. On a subset of M, the matrix A has a unique inverse A^{-1}, not necessarily bounded, given by

$$A^{-1} = \int_{-\infty}^{+\infty} \lambda^{-1} dE'(\lambda),$$

where $E'(\lambda) = E(\lambda)$ or $E(\lambda) - E(0)$ according as $\lambda < 0$ or $\lambda \geqq 0$. Finally, (5) is necessary and sufficient in order that y be in the domain of A^{-1}. Then (6) is $A^{-1}y$. Since (6) is orthogonal to all solutions of $Ax = 0$, it follows that (6) is the solution of (2) of least length.

If A is not self-adjoint, it is sufficient to apply the well-known device of replacing (2) by the equivalent system of equations

$$\begin{pmatrix} 0 & A \\ A^* & 0 \end{pmatrix} \begin{pmatrix} 0 \\ x \end{pmatrix} = \begin{pmatrix} y \\ 0 \end{pmatrix},$$

in which the operator represented by the 2-rowed matrix is self-adjoint (A^* denotes the adjoint of A). What thus results, in case of solvability, is again the solution of least length. Correspondingly, the above result can be thought of as a spectral formulation of the criteria of E. Schmidt [8], which he expresses in terms of limits of Gram determinants or equivalent "elementary" procedures.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] P. Hartman and A. Wintner, "An oscillation theorem for continuous spectra," *Proceedings of the National Academy of Sciences*, vol. 33 (1947), pp. 376-379.

[2] ——— and A. Wintner, "On the orientation of unilateral spectra," *American Journal of Mathematics*, vol. 70 (1948), pp. 309-316.

[3] ——— and A. Wintner, "Separation theorems for bounded Hermitian forms," *ibid.*, vol. 71 (1949), pp. 865-878.

[4] ——— and A. Wintner, "A separation theorem for continuous spectra," *ibid.*, vol. 71 (1949), pp. 650-663.

[5] E. L. Ince, "A proof of the impossibility of the coexistence of two Mathieu functions," *Proceedings of the Cambridge Philosophical Society*, vol. 21 (1922), pp. 117-120.

[6] ———, "The Mathieu equation with numerically large parameters," *Journal of the London Mathematical Society*, vol. 2 (1927), pp. 46-50.

[7] É. Picard, "Sur un théorème général relatif aux équations intégrales de première espèce," *Rendiconti del Circolo Matematico di Palermo*, vol. 29 (1910), pp. 79-96, § 1 (reprinted in Picard's *Selecta*, Paris (1928), pp. 206-211).

[8] E. Schmidt, "Ueber die Auflösung linearer Gleichungen mit unendlich vielen Unbekannten," *ibid.*, vol. 25 (1908), pp. 53-77.

[9] H. Weyl, "Ueber beschränkte quadratische Formen, deren Differenz vollstetig ist," *ibid.*, vol. 27 (1909), pp. 373-392.

[10] ———, "Ueber gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürlicher Funktionen," *Mathematische Annalen*, vol. 68 (1910), pp. 222-269.

AN APPLICATION OF THE THEORY OF BASIC SERIES TO THEOREMS OF BERNSTEIN-WIDDER TYPE.*

By PHILIP DAVIS.

1. Introduction. That the signs of the derivatives of a function of class C^∞ on an interval may influence its analytic character was first pointed out by S. Bernstein [1]. Bernstein's theorem may be stated as follows.

THEOREM 1. *Let $f(x)$ be a real valued function defined and of class C^∞ in $-1 < x < 1$. If each derivative of $f(x)$ is non-negative in $-1 < x < 1$, then $f(x)$ is necessarily analytic in that interval.*

Some years later, a result of a similar nature, but this time showing the effect of the signs of the even derivatives, was obtained by D. V. Widder [2].

THEOREM 2. *If $f(x)$ is real and of class C^∞ in $-1 < x < 1$ and if*

$$(1) \qquad (-1)^n f^{(2n)}(x) \geqq 0 \qquad (-1 < x < 1;\ n = 0, 1, 2, \cdots),$$

then $f(x)$ necessarily coincides with an entire function of exponential type at most π.

Functions satisfying the condition (1) have been termed completely convex. Subsequently Boas and Pólya [3] gave a generalization which contained as special cases both Theorems 1 and 2. In the present note these theorems are generalized in another direction by obtaining similar results for a number of familiar sets of differential operators. We shall find it convenient to formulate these results using the notion of the basic series of polynomials introduced by Whittaker [4]. Our work is essentially algebraic in character, inasmuch as all questions are referred back either to Theorem 1 or to Theorem 2.

2. Basic sets and series. In Whittaker's terminology, a set of polynomials $\{p_n(z)\}$ is called a basic set if an arbitrary polynomial can be expressed in a unique fashion as a finite linear combination of the p_n's. Whittaker has given the following necessary and sufficient condition that a set of polynomials be basic.

---

* Received April 6, 1950.

THEOREM 3. *Given the set of polynomials*

$$p_n(z) = \sum_{j=0}^{\infty} p_{nj} z^j, \qquad (n = 0, 1, 2, \cdots) \tag{2}$$

where in (2) *we have inserted dummy zero coefficients; a necessary and sufficient condition that the set* (2) *be basic is that the row-finite matrix*

$$P = (p_{nj}) \qquad (n, j = 0, 1, 2, \cdots) \tag{3}$$

have an inverse Q which is row-finite:

$$PQ = QP = I. \tag{4}$$

By a row-matrix is meant one in which each row has but a finite number of non-zero elements. It may also be remarked here that the multiplication of row-finite matrices is associative, so that the above condition is equivalent to the existence of a unique Q for which $QP = I$.[1]

With a given basic set of polynomials (2) with matrix $P = (p_{nj})$ and with unique row-finite inverse $Q = (q_{nj})$, there may be associated the following set of (possibly formal) power series:

$$q_n(z) = \sum_{k=0}^{\infty} q_{kn} z^k / k! \qquad (n = 0, 1, 2, \cdots). \tag{5}$$

The associated set may be characterized in the following way. The set $\{q_n(z)\}$ is associated to the basic set $\{p_n(z)\}$ of polynomials if and only if

$$\sum_{n=0}^{\infty} p_n(z) q_n(w) = e^{zw} \tag{6}$$

holds as a formal identity, in the sense that when substitutions are made, order of summation reversed, and coefficients compared, we have equality. it is easily seen that as a formal identity, (6) is equivalent to (4).

We shall give a second characterization of the associated set. For two power series $p(z) = \sum_{n=0}^{\infty} a_n z^n$, $q(z) = \sum_{n=0}^{\infty} b_n z^n$, introduce the inner product

$$[p, q] = \sum_{n=0}^{\infty} a_n b_n n!. \tag{7}$$

If either p or q is a polynomial, there can be no question of the convergence of the series in (7). The set (5) is associated to the basic set (2) if and only if the two sets are biorthonormal:

$$[p_i, q_j] = \delta_{ij}. \tag{8}$$

---

[1] See, e. g., Dienes [5].

With a given function f (of an appropriate class) and a given basic set, there may be associated the following formal expansion of f:

$$f(z) \sim \sum_{n=0}^{\infty} [q_n, f] p_n(z), \tag{9}$$

the so-called basic series for f. The study of the representability of functions in a series (9) forms a major chapter in the theory of functions. For our purposes, however, we need only know that equality in (9) holds whenever f is a polynomial. Since $\{p_n(z)\}$ is basic, there is a unique finite expansion of a polynomial f: $f = c_0p_0 + c_1p_1 + \cdots + c_kp_k$. Hence by (7) and (8), $c_i = [q_i, f]$.

We shall now list a number of basic sets together with their associated sets. Our list is intended to be suggestive rather than exhaustive, and may be augmented at will. Moreover, the examples have been chosen so that the associated functions are expressible in closed form. Except when otherwise noted, the 0-th polynomial is 1.

Maclaurin Polynomials:

$$p_n(z) = z^n/n\,!; \qquad q_n(z) = z^n. \tag{10}$$

Newton Interpolation Polynomials:

$$p_n(z) = (z - a_0)(z - a_1)\cdots(z - a_{n-1}), \quad a_i \text{ distinct};$$

$$q_n(z) = \begin{vmatrix} 1 & 1 & \cdots 1 \\ a_0 & a_1 & \cdots a_n \\ & & \cdots \\ a_0{}^{n-1} & a_1{}^{n-1} & \cdots a_n{}^{n-1} \\ e^{a_0z} & e^{a_1z} & \cdots e^{a_nz} \end{vmatrix} : \begin{vmatrix} 1 & 1 & \cdots 1 \\ a_0 & a_1 & \cdots a_n \\ & & \cdots \\ & & \\ a_0{}^n & a_1{}^n & \cdots a_n{}^n \end{vmatrix}. \tag{11}$$

Gontcharoff Polynomials (Whittaker [4]):

$$p_n(z) = \int_{a_0}^{z} dz_1 \int_{a_1}^{z_1} dz_2 \cdots \int_{a_{n-1}}^{z_{n-1}} dz_n, \qquad q_n(z) = z^n e^{a_n z}/n\,!. \tag{12}$$

Lidstone Polynomials (Widder [2]):

$$\begin{aligned} &p_0(z) = z, \qquad p_{2n+1}(z) = p_{2n}(1 - z); \\ &p_{2n-2}(z) = p_{2n}''(z), \qquad p_{2n}(0) = p_{2n}(1) = 0, \\ &q_{2n}(z) = z^{2n}, \qquad q_{2n+1,}z) = z^{2n}e^{z}. \end{aligned} \tag{13}$$

Hermite Polynomials:

$$(14)\quad p_n(z)=\sum_{k=0}^{[n/2]}(-1)^k(2z)^{n-2k}/k!(n-2k)!,\qquad q_n(z)=z^ne^{z^2/4}/2^nn!.$$

Appel Polynomials (Boas [6]):

$$A(w)e^{zw}=\sum_{n=0}^{\infty}p_n(z)w^n,\qquad q_n(z)=z^n/A(z),$$

$A(w)$ regular at $w=0$, $A(0)\neq 0$.

3. Applications.

We may now state our principal theorem.

THEOREM 4. *Let $\{p_n(z)\}$ be a basic set of polynomials with real coefficients, and let $\{q_n(z)\}$ be the (possibly formal) set of associated power series. For each n, let $[q_j(z),z^n]=q_{nj}$ have a constant sign as j varies. Then, with $f(x)$ a real valued function of class C^∞ on $-1<x<1$,*

$$(16)\qquad p_n(D)f(x)\geqq 0\qquad (n=0,1,2,\cdots;-1<x<1;D\equiv d/dx)$$

implies that $f(x)$ is analytic in $-1<x<1$.

Proof. We have, for some integer k_n,

$$z^n=\sum_{j=0}^{k_n}[q_j(z),z^n]p_j(z);\text{ hence }D^nf=\sum_{j=0}^{k_n}q_{nj}p_j(D)f.$$

The inequality (16) now implies that $D^nf(x)$ has constant sign in $-1<x<1$. The theorem is therefore an immediate consequence of Theorem 1.

It may happen that the associated series $q_n(z)$ have non-negative coefficients for all n. In other words, the basic set $\{p_n(z)\}$ has a positive inverse property. All the sets of our list, with proper restrictions on the arbitrary constants appearing therein, have this property.

COROLLARY 4.1. *If $a_i\geqq 0$, then the sets* (10), (11), (12), (13), (14) *have a positive inverse property. This is also true of the set* (15) *provided that $[A(w)]^{-1}$ has non-negative Maclaurin coefficients. For the polynomials of any of these sets, condition* (16) *implies analyticity in $-1<x<1$.*

Proof. Inspection of the corresponding q_n's reveals that under the above conditions their coefficients are indeed non-negative. For the set (11) we should add that $q_{nj}=\Sigma\, a_0^{i_0}a_1^{i_1}\cdots a_j^{i_j}$ where the summation is extended over all non-negative i's for which $i_0+i_1+\cdots+i_j=n-j$.

Theorem 4 is capable of further generalization, for Theorem 1, as stated, is but a special case of Bernstein's results. Actually, Bernstein has shown that if each successive difference of $f(x)$ is of constant sign in the interval, then, without assuming the existence of derivatives, it will follow that $f(x)$ is analytic in the interval. If therefore in Theorem 4, $p_n(D)$ be replaced by $p_n(\Delta)$, the conclusion follows without assuming the existence of derivatives.

By selecting the Newton Polynomials (11) and combining Theorems 4 and 2, we arrive at the following generalization of Widder's Theorem.

COROLLARY 4.2. *Let* $f(x)$ *be real and of class* C^∞ *on* $-1 < x < 1$, *and suppose that* a_i *are real. Then*

$$f(x) \geqq 0, \qquad (-1)^n(D^2+a_1^2)\cdots(D^2+a_n^2)f(x) \geqq 0 \tag{17}$$

$(n = 1, 2, \cdots; -1 < x < 1)$ *implies that* $f(x)$ *is completely convex on the interval and hence coincides with an entire function of exponential type at most* π.

HARVARD UNIVERSITY.

REFERENCES.

[1] S. Bernstein, "Sur la définition et les propriétés des fonctions analytiques d'une variable réelle," *Mathematische Annalen*, vol. 75 (1914), pp. 449-468.

[2] D. V. Widder, "Completely convex functions and Lidstone series," *Transactions of the American Mathematical Society*, vol. 51 (1942), pp. 387-398.

[3] R. P. Boas, Jr. and G. Pólya, "Influence of the signs of the derivatives of a function on its analytic character," *Duke Journal of Mathematics*, vol. 9 (1942), pp. 406-424.

[4] J. M. Whittaker, *Interpolatory Function Theory*, Cambridge Tracts No. 33, Cambridge, 1935.

[5] P. Dienes, "Notes on linear equations in infinite matrices," *Quarterly Journal of Mathematics*, vol. 3 (1932), pp. 253-268.

[6] R. P. Boas, Jr., "Exponential transforms and Appel polynomials," *Proceedings of the National Academy of Sciences*, vol. 34 (1948), pp. 481-483.

THE MEAN CONVERGENCE OF ORTHOGONAL SERIES.*

By G. Milton Wing.

---

1. Introduction. We consider a sequence of functions $\{\phi_n(x)\}$, $n = 0, 1, 2, \cdots$, orthonormal with respect to a measure

$$\int_a^b \phi_n(x)\phi_m(x)\,dF(x) = \delta_{nm},$$

and the corresponding formal expansion of an arbitrary function in terms of this sequence,

$$f(x) \sim \sum_{n=0}^{\infty} a_n\phi_n(x), \tag{1.1}$$

where $a_n = \int_a^b f(x)\phi_n(x)\,dF(x)$. While the local convergence properties of series of the type (1.1) have been thoroughly investigated for particular sets of $\phi_n(x)$, their mean convergence has been studied but little. This problem takes the form: for what values of p, $1 \leqq p < \infty$ does the existence of the integral $\int_a^b |f(x)|^p\,dF(x)$ imply that

$$\lim_{N\to\infty} \int_a^b |f(x) - \sum_{n=0}^{N} a_n\phi_n(x)|^p\,dF(x) = 0? \tag{1.2}$$

When equation (1.2) holds for every $f(x)$ such that $\int_a^b |f(x)|^p\,dF(x)$ exists we shall say that the sequence $\{\phi_n(x)\}$ forms a *basis* for the space of these functions.

In the case that $F(x)$ is an increasing, absolutely continuous function a different question may be posed. If $dF(x) = w(x)\,dx$ then the sequence $\{\phi_n(x)w^{\frac{1}{2}}(x)\}$ is orthonormal in the classical sense and one is led to the formal expansion

$$f(x) \sim \sum_{n=0}^{\infty} b_n\phi_n(x)w^{\frac{1}{2}}(x), \tag{1.3}$$

---

* Received November 16, 1949; revised April 2, 1950. Presented to the Society October 29, 1949. The research was carried on while the author held an Atomic Energy Commission Predoctoral Fellowship.

where $b_n = \int_a^b f(x)\phi_n(x)w^{\frac{1}{2}}(x)\,dx$. We now ask: for what values of p, $1 \leqq p < \infty$, does the measurability of $f(x)$ and the relation $\int_a^b |f(x)|^p\,dx < \infty$ (i. e., $f(x)$ in L^p) imply

$$\lim_{N\to\infty}\int_a^b \left| f(x) - \sum_{n=0}^{N} b_n\phi_n(x)w^{\frac{1}{2}}(x)\right|^p dx = 0?$$

Problems of this sort were first investigated by M. Riesz [15], who considered the case of trigonometric functions, and later by Schauder, Kober, Caton and Hille, and Paley [16, 8, 4, 10], who investigated other special sets of functions. Recently, H. Pollard has studied extensively the mean convergence of orthogonal polynomials [11, 12, 13, 14]. We shall make some additions to his results and apply similar methods to the problems associated with expansions in series of Bessel functions. It should be noted that the L^p theory of Hankel transforms (the integral analogue of Fourier-Bessel Series) has been studied by Busbridge [3].

2. General principles. Our interest will be confined to a sequence $\{\phi_n(x)\}$ orthonormal over a finite interval with respect to a non-negative weight function $w(x)$. We shall assume the elements of the sequence belong to L_w^p (i. e. $\int_a^b |\phi_n(x)|^p w(x)\,dx < \infty$ for all n) and to the conjugate space $L_w^{p'}$. We assume also that the sequence is closed over both of these spaces. The notation $\| f(x)\|_p$ will regularly be used to mean $\left\{\int_a^b |f(x)|^p\,dx\right\}^{1/p}$ and $S_N(f,x)$ will denote the N-th partial sum of the series in (1.1). The following principles will be used repeatedly.

(i) To establish that $\{\phi_n(x)\}$ is a basis for L_w^p, $p > 1$, it suffices to prove that, for all $f(x)$ in L_w^p,

$$\overline{\lim_{N\to\infty}} \| S_N(f,x)w^{1/p}(x)\|_p < +\infty.$$

The proof of this has been given by Pollard [13, p. 361].

(ii) If $\{\phi_n(x)\}$ is a basis for L_w^p, $p > 1$, then it is also a basis for $L_w^{p'}$.

This principle may be found in Banach [1, p. 108] for the case $w(x) \equiv 1$. His proof may be generalized to include our case.

CHAPTER I. BESSEL SERIES.

3. Orthogonality of sets of Bessel functions. It is well known [19, p. 134] that if $\{\lambda_n\}$ $(n = 1, 2, \cdots)$ is the sequence of successive positive roots of the equation

$$xJ_\nu'(x) + HJ_\nu(x) = 0, \tag{3.1}$$

where H is a real number and $\nu \geqq -1$, then

$$\int_0^1 xJ_\nu(\lambda_n x)J_\nu(\lambda_m x)\,dx = \tfrac{1}{2}\{(1 - \nu^2/\lambda^2{}_n)J_\nu{}^2(\lambda_n) + J_\nu'^2(\lambda_n)\}\delta_{nm}. \tag{3.2}$$

In case $H = +\infty$ the sequence, which we now denote by $\{\mu_n\}$, becomes the sequence of positive roots of $J_\nu(x)$. In place of (3.2) we have

$$\int_0^1 xJ_\nu(\mu_n x)J_\nu(\mu_m x)\,dx = \tfrac{1}{2}J_{\nu+1}{}^2(\mu_n)\,\delta_{nm}. \tag{3.3}$$

Series based upon the relation (3.3) are usually referred to as Fourier-Bessel series. They are somewhat easier to handle than the more general Dini series arising from equation (3.2). Fortunately, however, the convergence properties of the latter may be deduced from those of the simpler series.

4. Fourier-Bessel series of the type (1.3). We consider the sequence $\{(2x)^{\frac{1}{2}}J_\nu(\mu_n x)/J_{\nu+1}(\mu_n)\}$, and write

$$S_N(f, x) = \sum_{n=1}^{N} b_n(2x)^{\frac{1}{2}}J_\nu(\mu_n x)/J_{\nu+1}(\mu_n),$$

where $J_{\nu+1}(\mu_n)\,b_n = 2^{\frac{1}{2}}\int_0^1 t^{\frac{1}{2}}J_\nu(\mu_n t)f(t)\,dt.$

Then we have $S_N(f, x) = \int_0^1 (xt)^{\frac{1}{2}}U_N(t, x)f(t)\ dt$, where

$$U_N(t, x) = 2\sum_{n=1}^{N} J_\nu(\mu_n x)J_\nu(\mu_n t)/J_{\nu+1}{}^2(\mu_n).$$

The following lemma, due to MacRobert [9], is crucial in the proofs of many of the theorems of this chapter.

LEMMA 4.1. *If* $\nu \geqq -\frac{1}{2}$ *and* $A_N = (N + \frac{1}{2}\nu + \frac{1}{4})\pi$, *then*

$$|\,U_N(t, x) - \tfrac{1}{2}A_N\{J_\nu(A_N x)J_{\nu+1}(A_N t) - J_\nu(A_N t)J_{\nu+1}(A_N x)\}/(x - t)\,|$$
$$< C(xt)^{-\frac{1}{2}}\{(x + t)^{-1} + (2 - x - t)^{-1}\},$$

for all x *and all* t *in* $[0, 1]$.

This result we now use to obtain our first theorem concerning mean convergence of Bessel series.

THEOREM 4.1. *If* $\nu \geqq -\frac{1}{2}$ *and* $f(x)$ *is in* L^p *for* $p > 1$ *then*

$$\lim_{N\to\infty} \int_0^1 | f(x) - S_N(f,x) |^p \, dx = 0.$$

Remark. We note the relations $J_{\frac{1}{2}}(x) = (2/\pi x)^{\frac{1}{2}} \sin x$ and $J_{-\frac{1}{2}}(x) = (2/\pi x)^{\frac{1}{2}} \cos x$. Hence when $\nu = \pm \frac{1}{2}$ our series become trigonometric series, and our theorem therefore includes the classical result of M. Riesz on mean convergence of these series [15].

Proof of the theorem. Since the function $t^{-\nu} J_\nu(t)$ is entire [see 19, p. 40], the condition $\nu \geqq -\frac{1}{2}$ implies that the functions $t^{\frac{1}{2}} J_\nu(\mu_n t)$ are bounded in t in $[0,1]$. The coefficients a_n are therefore well-defined. The lemma suggests that we write

$$\begin{aligned} S_N(f,x) = \int_0^1 \{ & U_N(t,x) \\ & - \tfrac{1}{2} A_N [J_\nu(A_N x) J_{\nu+1}(A_N t) - J_\nu(A_N t) J_{\nu+1}(A_N x)]/(x-t) \\ & + \tfrac{1}{2} A_N [J_\nu(A_N x) J_{\nu+1}(A_N t) - J_\nu(A_N t) J_{\nu+1}(A_N x)]/(x-t) \} (xt)^{\frac{1}{2}} f(t) \, dt. \end{aligned}$$

From the lemma we deduce that

$$\begin{aligned} (4.1) \quad | S_N(f,x) | \leqq C & \int_0^1 \{ (x+t)^{-1} + (2-x-t)^{-1} \} | f(t) | \, dt \\ & + \tfrac{1}{2} \left| \int_0^1 (A_N x)^{\frac{1}{2}} J_\nu(A_N x) (A_N t)^{\frac{1}{2}} J_{\nu+1}(A_N t) f(t)/(x-t) \, dt \right| \\ & + \tfrac{1}{2} \left| \int_0^1 (A_N x)^{\frac{1}{2}} J_{\nu+1}(A_N x) (A_N t)^{\frac{1}{2}} J_\nu(A_N t) f(t)/(x-t) \, dt \right|. \end{aligned}$$

The convergence almost everywhere of the improper integrals, as Cauchy principal values, follows from the fact that the functions $(A_N t)^{\frac{1}{2}} J_\nu(A_N t) f(t)$ and $(A_N t)^{\frac{1}{2}} J_{\nu+1}(A_N t) f(t)$ are integrable [18, p. 132].

We shall now prove that

$$(4.2) \qquad \overline{\lim_{N\to\infty}} \, \| S_N(f,x) \|_p < +\infty.$$

By Hilbert's inequality [6, p. 226] the function $g(x) = \int_0^1 | f(t) | / (x+t) dt$

is in L^p so that $\| g(x) \|_p < \infty$. A simple change of variables allows us to make a similar statement concerning $\int_0^1 | f(t) | / (2 - x - t)\, dt$.

The remaining two integrals of (4. 1) are similar, and we need examine only one of them. We follow the method used by M. Riesz in his treatment of trigonometric series [15]. Write

$$I_N(x) = \int_0^1 (A_N x)^{\frac{1}{2}} J_\nu(A_N x)(A_N t)^{\frac{1}{2}} J_{\nu+1}(A_N t) f(t)/(x - t)\, dt$$
$$= B_N(x) \int_0^1 C_N(t) f(t)/(x - t)\, dt,$$

where $B_N(x) = (A_N x)^{\frac{1}{2}} J_\nu(A_N x)$, $C_N(t) = (A_N t)^{\frac{1}{2}} J_{\nu+1}(A_N t)$.

From the asymptotic form of the Bessel functions [19, p. 194], we conclude that $B_N(x)$ and $C_N(t)$ are bounded in N, x, and t. Hence $C_N(t) f(t)$ is a member of L^p. It follows by a theorem of Riesz [15] that $I_N(x)$ is also in L^p and $\| I_N(x) \|_p < A \| C_N(x) f(x) \|_p < A \| f(x) \|_p < A$, since B_N and C_N are bounded.

We now apply Minkowski's inequality to (4. 1) to get

$$\| S_N(f, x) \|_p \leqq C \left\| \int_0^1 \{(x + t)^{-1} + (2 - x - t)^{-1}\} | f(t) |\, dt \right\|_p$$
$$+ \tfrac{1}{2} \left\| B_N(x) \int_0^1 C_N(t) f(t)/(x - t)\, dt \right\|_p$$
$$+ \tfrac{1}{2} \left\| C_N(x) \int_0^1 B_N(t) f(t)/(x - t)\, dt \right\|_p < + \infty.$$

Hence (4. 2) is established.

If it can now be shown that the sequence $\{(2x)^{\frac{1}{2}} J_\nu(\mu_n x)/J_{\nu+1}(\mu_n)\}$ is closed in L^p, principle (i) of § 2 may be applied to complete the proof. W. H. Young has proved if $t^{\frac{1}{2}} f(t)$ is in L, then the vanishing of $\int_0^1 t f(t) J_\nu(\mu_n t)\, dt$ for all n implies that $f(t) \equiv 0$ [21, pp. 174-175]. From this follows the completeness of our sequence in L^p, and hence its closure [see 7, p. 200].

5. Fourier-Bessel series of the type (1. 1). We turn next to series of type (1. 1) with $\phi_n(x) = 2^{\frac{1}{2}} J_\nu(\mu_n x)/J_{\nu+1}(\mu_n)$ and

$$J_{\nu+1}(\mu_n) a_n = 2^{\frac{1}{2}} \int_0^1 t f(t) J_\nu(\mu_n t)\, dt.$$

The partial sums have the form $T_N(f,x) = \int_0^1 tf(t)\,U_N(t,x)\,dt$. We prove the following theorem.

THEOREM 5.1. *If* $\nu \geqq -\frac{1}{2}$ *and* $f(t)$ *is in* L_t^p, $4/3 < p < 4$, *then*

$$\lim_{N\to\infty} \int_0^1 | f(x) - T_N(f,x) |^p x\, dx = 0.$$

Proof. The coefficients a_n exist provided $t^{\frac{1}{2}} | f(t) |$ is in L, since $t^{\frac{1}{2}} J_\nu(\mu_n t)$ is bounded in t in $[0,1]$. Now

$$\int_0^1 t^{\frac{1}{2}} | f(t) | \,dt \leqq \left\{ \int_0^1 t^{(\frac{1}{2}-1/p)p'} dt \right\}^{1/p'} \left\{ \int_0^1 t \,| f(t) |^p dt \right\}^{1/p}.$$

The last integral exists by assumption. The first on the right converges provided $(\frac{1}{2} - 1/p)p' = (1/p' - \frac{1}{2})p' > -1$ or $p' < 4$, as assumed.

Proceeding exactly as in Theorem 4.1, we find that

$$(5.1)\quad | T_N(f,x) | \leqq Cx^{-\frac{1}{2}} \int_0^1 t^{\frac{1}{2}} | f(t) | \{(x+t)^{-1} + (2-x-t)^{-1}\} dt$$
$$+ \tfrac{1}{2} \left| \int_0^1 A_N J_\nu(A_N x) J_{\nu+1}(A_N t)\, t f(t) / (x-t)\, dt \right|$$
$$+ \tfrac{1}{2} \left| \int_0^1 A_N J_\nu(A_N t) J_{\nu+1}(A_N x)\, t f(t) / (x-t)\, dt \right|.$$

This expression must now be treated more carefully than the similar one for $| S_N(f,x) |$ encountered in § 4. Our goal is to prove

$$(5.2)\quad \overline{\lim_{N\to\infty}} \| T_N(f,x) x^{1/p} \|_p < +\infty.$$

We begin by considering

$$(5.3)\quad g(x) = x^{1/p-\frac{1}{2}} \int_0^1 t^{\frac{1}{2}} | f(t) | / (x+t)\, dt$$
$$= \int_0^1 t^{1/p} | f(t) | / (x+t) dt + \int_0^1 t^{1/p} | f(t) | \{(t/x)^{\frac{1}{2}-1/p} - 1\} / (x+t) dt$$

$= I_1(x) + I_2(x)$. Hilbert's inequality again applies to $I_1(x)$. By the converse of Hölder's inequality [6, p. 120], $I_2(x)$ belongs to L^p if the double integral

$$(5.4) \qquad I = \int_0^1 h(x)\,dx \int_0^1 t^{1/p}\,|f(t)|\{(t/x)^{\frac{1}{2}-1/p} - 1\}/(x+t)\,dt$$

is bounded for every $h(x)$ in $L^{p'}$.

Applying Hölder's inequality for double integrals to (5.4) we have

$$|I| = \left|\int_0^1 dx \int_0^1 h(x)\{[(t/x)^{\frac{1}{2}-1/p} - 1]/(x+t)\}^{1/p'} t^{1/p}\,|f(t)|\,(t/x)^{-1/pp'}\right.$$

$$\left.\cdot\{[t/x)^{\frac{1}{2}-1/p} - 1]/(x+t)\}^{1/p}(t/x)^{1/pp'}\,dt\right|$$

$$\leqq \left\{\int_0^1 |h(x)|^{p'}dx \int_0^1 |(t/x)^{\frac{1}{2}-1/p} - 1\,|\,(t/x)^{-1/p}/(x+t)\,dt\right\}^{1/p'}$$

$$\cdot\left\{\int_0^1 t\,|f(t)|^p dt \int_0^1 |(t/x)^{\frac{1}{2}-1/p} - 1\,|\,(t/x)^{1/p'}/(x+t)\,dx\right\}^{1/p}.$$

In view of the conditions imposed on p, it now suffices to prove the boundedness, for all x in $(0,1)$, of

$$J_1 = \int_0^1 |(t/x)^a - (t/x)^b\,|/(x+t)\,dt,$$

where $0 > a > -1$, $0 > b > -1$. The trivial substitution $z = tx$ yields

$$J_1 = \int_0^{1/x} |\,z^a - z^b\,|/(1+z)\,dz \leqq \int_0^\infty |\,z^a - z^b\,|/(1+z)\,dz.$$

The last integral is clearly finite. Hence I converges and $g(x)$ is in L^p. Similarly, $g_1(x) = x^{1/p-\frac{1}{2}} \int_0^1 t^{\frac{1}{2}}\,|f(t)|/(2-x-t)\,dt$ is a member of L^p.

Finally we turn to the last two integrals of (5.1). It suffices to consider just one of these. Let

$$G_N(x) = x^{1/p} \int_0^1 A_N J_\nu(A_N x) J_{\nu+1}(A_N t)\,t f(t)/(x-t)\,dt$$

$$= x^{1/p-\frac{1}{2}} B_N(x) \int_0^1 C_N(t)\,t^{\frac{1}{2}} f(t)/(x-t)\,dt.$$

In proving that $G_N(x)$ is in L^p, it will do no harm to suppose $f(t) \geqq 0$. Write

$$G_N(x) = B_N(x) \int_0^1 C_N(t)\,t^{1/p} f(t)/(x-t)\,dt$$

$$+ B_N(x) \int_0^1 C_N(t)\,t^{1/p} f(t)\{(t/x)^{\frac{1}{2}-1/p} - 1\}/(x-t)\,dt.$$

The first integral is covered by Riesz's theorem. The second may be treated by the device used in investigating $I_2(x)$ of (5.3) and leads to

$$J_2 = \int_0^1 |(t/x)^a - (t/x)^b|/(x-t)\,dt.$$

This is finite for x in $(0,1)$. Hence

$$\| G_N(x) \|_p \leqq A \left\| \int_0^1 C_N(t) t^{1/p} f(t)/(x-t)\,dt \right\|_p + A \left\| \int_0^1 C_N(t) t^{1/p} f(t) \{ (t/x)^{\frac{1}{2}-1/p} - 1 \}/(x-t)\,dt \right\|_p \leqq A.$$

Here A is independent of N since the B's are bounded in N. This establishes (5.2).

We need now only establish the closure of the sequence $\{J_\nu(\mu_n t)\}$ in L_t^p. Since $f(t)$ is in L_t^p for $p > 4/3$ it follows that $t^{\frac{1}{2}} f(t)$ is in L. Hence we may again use the result of Young and assert that the sequence is complete in L_t^p for $p > 4/3$, and so closed there for $p < 4$.

6. A counterexample. We have seen that the condition $p > 4/3$ is necessary to insure the existence of the coefficients a_n, provided we allow the order ν of the Bessel functions to be as small as $-\frac{1}{2}$. Our argument depended upon the fact that $t^{\frac{1}{2}} J_\nu(\mu_n t)$ is bounded in t, $0 \leqq t \leqq 1$, for $\nu \geqq -\frac{1}{2}$. Since, more generally, $t^{-\nu} J_\nu(\mu_n t)$ is bounded in t, one might expect that the conditions on p should really depend upon ν. The following example shows that this is not the case. In view of principle (ii) of § 2, we may confine our attention to the interval $1 \leqq p < 4/3$.

THEOREM 6.1. *There exists a function belonging to L_t^p, $1 \leqq p < 4/3$, whose Fourier-Bessel series does not converge in L_t^p.*

Proof. The function $f(t) = t^{-3/2}$ belongs to L_t^p for $p < 4/3$. To prove that its Fourier-Bessel series does not converge in the mean it suffices to prove that

$$\overline{\lim_{n \to \infty}} |a_n| \left\{ \int_0^1 t\, |J_\nu(\mu_n t)|^p\, dt \right\}^{1/p} > 0,$$

where

$$J_{\nu+1}^2(\mu_n) a_n = 2 \int_0^1 t^{-\frac{1}{2}} J_\nu(\mu_n t)\,dt. \tag{6.1}$$

We may confine our attention to the case $\nu > -\frac{1}{2}$ since the coefficients a_n

do not even exist when $\nu = -\frac{1}{2}$. Substituting z for $t\mu_n$ in (6.1) we have [see 20, p. 383]

$$(6.2) \qquad \lim_{n\to\infty} a_n \mu_n^{\frac{1}{2}} J_{\nu+1}{}^2(\mu_n) = 2 \int_0^\infty z^{-\frac{1}{2}} J_\nu(z)\,dz = 2^{-\frac{1}{2}} \Gamma(\tfrac{1}{2}\nu + \tfrac{1}{4}) / \Gamma(\tfrac{1}{2}\nu + \tfrac{3}{4})$$

$= A > 0$. Now

$$(6.3) \qquad \int_0^1 t\,|J_\nu(\mu_n t)|^p\,dt = \mu_n{}^{-2} \int_0^{\mu_n} z\,|J_\nu(z)|^p\,dz > B_p \mu_n{}^{-p/2},$$

as one may verify by use of the asymptotic formula

$$J_\nu(z) \sim (2/\pi z)^{\frac{1}{2}} \cos(z - \tfrac{1}{2}\nu\pi - \tfrac{1}{4}\pi)$$

[19, p. 195]. Combining (6.2) and (6.3) we get

$$\overline{\lim_{n\to\infty}}\,|a_n| \left\{ \int_0^1 t\,|J_\nu(\mu_n t)|^p\,dt \right\}^{1/p} > \overline{\lim_{n\to\infty}}\, 2AB_p/\mu_n J_{\nu+1}{}^2(\mu_n) > 0,$$

as desired.

It is interesting to note that the interval of mean convergence, $4/3 < p < 4$, is the same as that for Legendre series and for series of more general types of polynomials to be discussed in § 9. In all of these cases the question of convergence at the endpoints of the interval is still open.

7. Dini series. From (3.2) it is clear that the functions $k_n x^{\frac{1}{2}} J_\nu(\lambda_n x)$ where $2/k_n{}^2 = (1 - \nu^2/\lambda_n{}^2) J_\nu{}^2(\lambda_n) + J_\nu'^2(\lambda_n)$ form an orthonormal sequence over $[0, 1]$ when $\nu \geqq -1$. We are led to consider the formal expansion

$$(7.1) \qquad f(x) \sim \sum_{n=0}^\infty k_n \beta_n x^{\frac{1}{2}} J_\nu(\lambda_n x),$$

where $\beta_n = k_n \int_0^1 f(t) t^{\frac{1}{2}} J_\nu(\lambda_n t)\,dt$. Dini has shown in his study of pointwise convergence of such series that (7.1) is not always the correct formal expansion [5]. Rather, one must add to the series of (7.1) a term $B_0(x)$ where

$$B_0(x) = 0 \text{ if } H + \nu > 0;\ B_0(x) = 2(\nu + 1) x^{\nu+\frac{1}{2}} \int_0^1 t^{\nu+\frac{1}{2}} f(t)\,dt \text{ if } H + \nu = 0;$$

$$B_0(x) = 2\lambda_0{}^2 I_\nu(\lambda_0 x) x^{\frac{1}{2}} / \{(\lambda_0{}^2 + \nu^2) I_\nu{}^2(\lambda_0) - \lambda_0{}^2 I_\nu'^2(\lambda_0)\} \int_0^1 t^{\frac{1}{2}} f(t) I_\nu(\lambda_0 t)\,dt \quad \text{if } H + \nu < 0.$$

Here $\pm i\lambda_0$ are the two purely imaginary zeros of (3.1) known to exist when

$H + \nu < 0$ and $I_\nu(x) = e^{-\frac{1}{2}\nu\pi i}J_\nu(xe^{\frac{1}{2}\pi i})$. Denoting the partial sums of the amended series by $\sigma_N(f, x)$, we have $\sigma_N(f, x) = \int_0^1 (xt)^{\frac{1}{2}}C_N(t, x)f(t)\,dt$, where

$$C_N(t, x) = A_0(x, t) + \sum_{n=1}^{N} k_n^2 J_\nu(\lambda_n x)J_\nu(\lambda_n t),$$

$$A_0(x, t) = 0 \text{ if } H + \nu > 0;\ A_0(x, t) = 2(\nu + 1)x^\nu t^\nu \text{ if } H + \nu = 0;$$

$$A_0(x, t) = 2\lambda_0^2 I_\nu(\lambda_0 x)I_\nu(\lambda_0 t)/\{(\lambda_0^2 + \nu^2)I_\nu^2(\lambda_0) - \lambda_0^2 I_\nu'^2(\lambda_0)\} \quad \text{if } H + \nu < 0.$$

We now prove the following theorem.

Theorem 7.1. *If* $\nu \geqq -\frac{1}{2}$ *and* $f(x)$ *is in* L^p, $p > 1$, *then*

$$\lim_{N\to\infty} \int_0^1 |f(x) - \sigma_N(f, x)|^p\,dx = 0.$$

Proof. Using the notation of the previous section we let μ_m be the largest zero of $J_\nu(x)$ which is less than λ_{N+1}. Then it may be shown that

$$|C_N(t, x) - U_m(t, x)| < C/(xt)^{\frac{1}{2}}(2 - x - t) \text{ for all } x \text{ and } t \text{ in } [0, 1]$$

[19, p. 599]. Hence

$$|\sigma_N(f, x)| = \left| \int_0^1 (xt)^{\frac{1}{2}}f(t)\{C_N(t, x) - U_m(t, x) + U_m(t, x)\}dt \right|$$
$$\leqq C \int_0^1 |f(t)|/(2 - x - t)\,dt + |S_m(f, x)|.$$

The arguments of § 4 yield

$$\| \sigma_N(f, x) \|_p < A_p + \| S_m(f, x) \|_p < A_p.$$

To complete the proof we must establish the closure in L^p, $p > 1$, of the orthogonal sequence $\{\phi_n(x)\}$, where $\phi_n(x) = x^{\frac{1}{2}}J_\nu(\lambda_n x)$ for $n > 0$, and $\phi_0(x) = 0$, $x^{\nu+\frac{1}{2}}$ or $x^{\frac{1}{2}}I_\nu(\lambda_0 x)$, according as $H + \nu > 0$, $H + \nu = 0$, or $H + \nu < 0$. Young has proved the completeness of $\{\phi_n(x)\}$ and the closure then follows. This establishes the theorem.

To complete this section we write $\alpha_n = k_n \int_0^1 tf(t)J_\nu(\lambda_n t)\,dt$, and consider $\tau_N(f, x) = C_0(x) + \sum_{n=1}^{N} \alpha_n k_n J_\nu(\lambda_n x)$, where

$C_0(x) = 0$ if $H + \nu > 0$; $C_0(x) = 2(\nu + 1)x^\nu \int_0^1 t^{\nu+1} f(t)\,dt$ if $H + \nu = 0$;

$$C_0(x) = 2\lambda_0^2 I_\nu(\lambda_0 x)/\{(\lambda_0^2 + \nu^2) I_\nu^2(\lambda_0) - \lambda_0^2 I_\nu'^2(\lambda_0)\} \int_0^1 t f(t) I_\nu(\lambda_0 t)\,dt$$

if $H + \nu < 0$.

The methods we have employed in this paper now suffice to prove the following theorem.

THEOREM 7.2. *If* $\nu \geqq -\frac{1}{2}$ *and* $f(t)$ *is in* L_t^p, $4/3 < p < 4$, *then*

$$\lim_{N\to\infty} \int_0^1 |f(x) - \tau_N(f, x)|^p x\,dx = 0. \tag{7.2}$$

Moreover, there exists a function in L_t^p, $1 \leqq p < 4/3$ (*or* $p > 4$) *such that* (7.2) *fails to hold.*

CHAPTER II. ORTHOGONAL POLYNOMIALS.

8. In this chapter we shall extend some of the results of H. Pollard on the problem of mean convergence of series or orthogonal polynomials. The first theorem, without proof, has been announced in a special case by Pollard [11]. The second improves one of his results, while the last make it possible to absorb a wide class of polynomials into the existing theory.

9. Extension of a theorem of Pollard. The following theorem is due to Pollard in the special case when $w(x) = (1 - x)^\alpha (1 + x)^\beta$, $\alpha \geqq -\frac{1}{2}$, $\beta \geqq -\frac{1}{2}$.

THEOREM 9.2. *Let the polynomials* $\{p_n(x)\}$ *be orthonormal on* $[-1, 1]$ *with weight function* $w(x)$. *Let* $(1 - x^2)^{-1} q_n(x)$ *denote the polynomials orthonormal with respect to the weight function* $w(x)(1 - x^2)$ *and let the following conditions hold* [see 13, p. 356]:

(H-1) $w(x) \geqq 0$ *almost everywhere on* $[-1, 1]$;

(H-2) $\int_{-1}^1 |\log w(x)|\,(1 - x^2)^{-\frac{1}{2}}\,dx < \infty$;

(H-3) $(1 - x^2)^{1/4}\,|p_n(x)|\,w^{\frac{1}{2}}(x) \leqq A$ *for all* x *in* $[-1, 1]$ *and all* $n = 0, 1, 2, \cdots$;

(H-4) $(1 - x^2)^{-1/4}\,|q_n(x)|\,w^{\frac{1}{2}}(x) \leqq A$ *for all* x *in* $[-1, 1]$ *and all* $n = 0, 1, 2, \cdots$.

Then, if $f(t)$ *is in* L^p, $4/3 < p < 4$, *and* $b_n = \int_{-1}^{1} f(x) p_n(x) w^{\frac{1}{2}}(x)\, dx$, *we have*

$$\lim_{N\to\infty} \int_{-1}^{1} \left| f(x) - \sum_{n=0}^{N} b_n w^{\frac{1}{2}}(x) p_n(x) \right|^p dx = 0. \tag{9.1}$$

Proof. We first show that the coefficients b_n are uniformly bounded. For, using (H-3),

$$\begin{aligned} | b_n | = \left| \int_{-1}^{1} f(x) p_n(x) w^{\frac{1}{2}}(x)\, dx \right| &\leqq \left\{ \int_{-1}^{1} | f(x) |^p\, dx \right\}^{1/p} \\ &\quad \cdot \left\{ \int_{-1}^{1} | p_n(x) w^{\frac{1}{2}}(x) |^{p'}\, dx \right\}^{1/p'} \end{aligned}$$

$$\leqq A \, \| f(x) \|_p \left\{ \int_{-1}^{1} (1 - x^2)^{-p'/4} dx \right\}^{1/p'} < A_p, \text{ since } p' < 4.$$

We now write $S_N(f, x) = \sum_{n=0}^{N} b_n w^{\frac{1}{2}}(x) p_n(x)$, and seek to prove that $\overline{\lim}_{N\to\infty} \| S_N(f, x) \|_p < \infty$. Substituting the value of b_n and rearranging somewhat, we have

$$S_N(f, x) = w^{\frac{1}{2}}(x) \int_{-1}^{1} f(t) w^{\frac{1}{2}}(t) k_N(x, t)\, dt.$$

Here [see 13, p. 358].

$$\begin{aligned} k_N(x, t) = \alpha_N p_{N+1}(x)\, q_N(t)/(x - t) &+ \beta_N q_N(x) p_{N+1}(t)/(x - t) \\ &+ \gamma_N p_{N+1}(x) p_{N+1}(t), \end{aligned}$$

the constants α_N, β_N, γ_N being bounded in N. Therefore

$$\begin{aligned} S_N(f, x) = \alpha_N p_{N+1}(x) w^{\frac{1}{2}}(x) &\int_{-1}^{1} f(t) w^{\frac{1}{2}}(t) q_N(t)/(x - t)\, dt \\ + \beta_N q_N(x) w^{\frac{1}{2}}(x) &\int_{-1}^{1} f(t) w^{\frac{1}{2}}(t) p_{N+1}(t)/(x - t)\, dt \\ + \gamma_N p_{N+1}(x) w^{\frac{1}{2}}(x) &\int_{-1}^{1} f(t) w^{\frac{1}{2}}(t) p_{N+1}(t)\, dt. \end{aligned}$$

The last integral is just b_{N+1} and so is bounded. (H-3) and (H-4) then yield

$$\begin{aligned} \| S_N(f, x) \|_p < A &\left\| (1 - x^2)^{-\frac{1}{4}} \int_{-1}^{1} f(t) w^{\frac{1}{2}}(t) q_N(t)/(x - t)\, dt \right\|_p \\ + A &\left\| (1 - x^2)^{\frac{1}{4}} \int_{-1}^{1} f(t) w^{\frac{1}{2}}(t) p_{N+1}(t)/(x - t)\, dt \right\|_p + A. \end{aligned} \tag{9.2}$$

We now write $q_N(t)w^{\frac{1}{2}}(t) = A_N(t)(1-t^2)^{\frac{1}{4}}$, where $A_N(t)$ is bounded in N and t. Hence

$$\left\| (1-x^2)^{-\frac{1}{4}} \int_{-1}^{1} f(t)w^{\frac{1}{2}}(t)q_N(t)/(x-t)\,dt \right\|_p$$
$$\leqq \left\| \int_{-1}^{1} A_N(t)f(t)/(x-t)\,dt \right\|_p$$
$$+ \left\| \int_{-1}^{1} A_N(t)f(t)\{(1-t^2)^{\frac{1}{4}}(1-x^2)^{-\frac{1}{4}}-1\}/(x-t)\,dt \right\|_p.$$

According to the theorem of M. Riesz [15] the first term is bounded. To settle the second we need only prove that

$$\int_{-1}^{1} |f(t)| \left| (1-t^2)^{\frac{1}{4}}(1-x^2)^{-\frac{1}{4}}-1 \right| / |x-t|\,dt$$

is in L^p. This follows at once from a result of Pollard [13, p. 363]. The second integral of (9.2) is handled similarly.

The proof is now completed by noting that the sequence $\{w^{\frac{1}{2}}(x)p_n(x)\}$ is closed in L^p.

10. Weight functions of class B.

We shall say that the weight function $w(x) = t(x)(1-x)^\alpha(1+x)^\beta$ $(\alpha \geqq -\frac{1}{2}, \beta \geqq -\frac{1}{2})$ belongs to the class **B** if $t(x)$ is positive on $[-1,1]$ and has there a continuous derivative with the property that $t'(x+h) - t'(x) = O(n^{-2}|h|)$ as $h \to 0$ uniformly in $[-1,1]$.

The class **B** has been studied extensively by S. Bernstein. It may be shown that conditions (H-1) thru (H-4) of Theorem 9.1 are satisfied for a sequence orthonormal with respect to a member of **B** [2, pp. 267-268; 13, p. 362]. The case $t(x) \equiv 1$ yields the Jacobi polynomials $P_n^{(\alpha,\beta)}(x)$ [17, p. 57]. We now prove the following theorem concerning mean convergence of Jacobi series.

THEOREM 10.1. *Let* $w(x) = (1-x)^\alpha(1+x)^\beta$ $(\alpha \geqq -\frac{1}{2}, \beta \geqq -\frac{1}{2})$. *Then, if* $f(x)$ *is in* L^p, $4/3 < p < 4$, *equation* (9.1) *holds. Further, for* $1 \leqq p < 4/3$ (*or* $p > 4$), *there exists a function in* L^p *for which* (9.1) *fails to hold.*

Proof. The first part of the theorem follows at once from Theorem 9.1. For the second part it suffices to find a function $f(x)$ in L^p, $1 \leqq p < 4/3$, such that

$$\overline{\lim_{n\to\infty}} \mid b_n \mid \left\{ \int_{-1}^{1} \mid (1-x)^{\frac12\alpha}(1+x)^{\frac12\beta} P_n^{(\alpha,\beta)}(x) \mid^p dx \right\}^{1/p} > 0.$$

Here [see 17, p. 238]

$$b_n h_n^{(\alpha,\beta)} = \int_{-1}^{1} f(x)\,(1-x)^{\frac12\alpha}(1+x)^{\frac12\beta} P_n^{(\alpha,\beta)}(x)\,dx.$$

The function $f(x) = (1+x)^{\frac12\beta}(1-x)^{-3/4}$ is in L^p and [see 17, p. 249],

$$b_n h_n^{(\alpha,\beta)} = \int_{-1}^{1} (1+x)^{\beta}(1-x)^{\frac12\alpha-3/4} P_n^{(\alpha,\beta)}(x)\,dx \sim n^{\frac12}. \tag{10. 1}$$

Further, $I = \int_0^1 (1-x)^{\frac12\alpha} \mid P_n^{(\alpha,\beta)}(x) \mid dx$

$$= \int_0^1 (1-x)^{\frac12\alpha}(1+x)^{\frac12\beta}(1+x)^{-\frac12\beta} \mid P_n^{(\alpha,\beta)}(x) \mid dx$$

$$\leqq \left\{ \int_{-1}^{1} \mid (1-x)^{\frac12\alpha}(1+x)^{\frac12\beta} P_n^{(\alpha,\beta)}(x) \mid^p dx \right\}^{1/p} \left\{ \int_0^1 (1+x)^{-\frac12\beta p'}\,dx \right\}^{1/p'}.$$

Now $I \sim n^{-\frac12}$ [17, p. 168], and the last integral is finite. Therefore, we may select an $A > 0$ such that

$$\left\{ \int_{-1}^{1} \mid (1-x)^{\frac12\alpha}(1+x)^{\frac12\beta} P_n^{(\alpha,\beta)}(x) \mid^p dx \right\}^{1/p} > A n^{-\frac12} \tag{10. 2}$$

for all n. From (10. 1) and (10. 2) we obtain the desired result.

THEOREM 10. 2. *Let $w(x) \varepsilon$ B. Let $f(x) \varepsilon L_w{}^p$ for a value of p satisfying*

$$4 \max \left\{ \frac{\alpha+1}{2\alpha+3}, \frac{\beta+1}{2\beta+3} \right\} < p < 4 \min \left\{ \frac{\alpha+1}{2\alpha+1}, \frac{\beta+1}{2\beta+1} \right\},$$

and define $a_n = \int_{-1}^{1} f(x) p_n(x) w(x)\,dx$. Then

$$\lim_{N\to\infty} \int_{-1}^{1} \mid f(x) - \sum_{n=0}^{N} a_n p_n(x) \mid^p w(x)\,dx = 0.$$

Proof. According to a result of Pollard [13, p. 363], it suffices to prove that the kernels

$$K_{\pm}(x,t) = \left| \left\{ \left(\frac{1-t^2}{1-x^2} \right)^{\pm\frac14} \left(\frac{w(t)}{w(x)} \right)^{\frac12-1/p} - 1 \right\} / (x-t) \right| \tag{10. 3}$$

have the property that $\int_{-1}^{1} K_{\pm}(x,t)\,g(t)\,dt$ belong to L^p whenever $g(t)$ does.

Our method of proof is an extension of one used already by him [14]. We shall deal only with K_+; a similar treatment may be used for K_-.

It suffices to show [14] that

$$F(x) = \int_{-1}^{1} K_+(x,t) \left(\frac{1-x^2}{1-t^2}\right)^{1/p} dt =$$
$$\int_{-1}^{1} \left| \left\{ \left(\frac{1-x}{1-t}\right)^a \left(\frac{1+x}{1+t}\right)^b \frac{T(t)}{T(x)} - \left(\frac{1-x}{1-t}\right)^c \left(\frac{1+x}{1+t}\right)^d \right\} \frac{1}{(x-t)} \right| dt$$

is bounded in $(-1,1)$, where

$$a = -\tfrac{1}{4} + \alpha(1/p - \tfrac{1}{2}) + 1/p, \qquad b = -\tfrac{1}{4} + \beta(1/p - \tfrac{1}{2}) + 1/p,$$
$$c = d = 1/p$$

and $T(x) = \{t(x)\}^{\frac{1}{2}-1/p}$. We remark that a, b, c, and d all lie in the open interval $(0,1)$.

Assume for the present that $0 \leqq x < 1$ and write $F(x)$ in the form

$$F(x) = \int_{-1}^{-\frac{1}{2}(1+x)} + \int_{-\frac{1}{2}(1+x)}^{x} + \int_{x}^{\frac{1}{2}(1+x)} + \int_{\frac{1}{2}(1+x)}^{1}$$

$= A(x) + B(x) + C(x) + D(x)$. (This decomposition is attributed by Pollard to P. Erdös.) The last integral is easily dealt with. Since the conditions on $t(x)$ imply that $T(x)$ is bounded above and below, $0 < m \leqq T(x) \leqq M < \infty$, we may write

$$D(x) \leqq K\{[(1-x)^a(1+x)^b/T(x)] \int_{\frac{1}{2}(1+x)}^{1} (1-t)^{-a} \,|\, x-t \,|^{-1} dt$$
$$+ (1-x)^c(1+x)^d \int_{\frac{1}{2}(1+x)}^{1} (1-t)^{-c} \,|\, x-t \,|^{-1} dt\}$$
$$\leqq K\{[(1-x)^a(1+x)^b/\tfrac{1}{2}m(1-x)] \int_{\frac{1}{2}(1+x)}^{1} (1-t)^{-a} dt$$
$$+ [(1-x)^c(1+x)^d/\tfrac{1}{2}(1-x)] \int_{\frac{1}{2}(1+x)}^{1} (1-t)^{-c} dt\} \leqq K.$$

The third integral is more difficult. Writing u for $t - x$, we get

$$C(x) = \int_{0}^{\frac{1}{2}(1-x)} u^{-1} \left| \left(\frac{1-x}{1-u-x}\right)^a \left(\frac{1+x}{1+u+x}\right)^b \frac{T(u+x)}{T(x)} - \left(\frac{1-x}{1-u-x}\right)^c \left(\frac{1+x}{1+u+x}\right)^d \right| du. \tag{10.4}$$

By the mean value theorem,

$$T(u+x)/T(x) = 1 + uT'(x+\theta_x u)/T(x) \qquad (0 < \theta_x < 1)$$

for all x and u such that $-1 \leqq x \leqq 1$, $-1 \leqq u + x \leqq 1$. Formula (10.4) then gives us

$$\begin{aligned} C(x) \leqq \int_0^{\frac{1}{2}(1-x)} u^{-1} \Bigg| & \left(1+\frac{u}{1-u-x}\right)^a \left(1-\frac{u}{1+u+x}\right)^b \\ & -\left(1+\frac{u}{1-u-x}\right)^c \left(1-\frac{u}{1+u+x}\right)^d \Bigg| \, du \\ + \int_0^{\frac{1}{2}(1-x)} \Bigg| & \frac{T'(x+\theta_x u)}{T(x)} \left(1+\frac{u}{1-u-x}\right)^a \left(1-\frac{u}{1+u+x}\right)^b \Bigg| \, du. \end{aligned}$$

The first integral is known to be bounded [14]. In the second we observe that the conditions imposed upon $t(x)$ imply that $T'(x+\theta_x u)$ is continuous in $x+\theta_x u$, and hence bounded. Further, since $0 < a < 1$,

$$\left(1+\frac{u}{1-u-x}\right)^a \leqq 1 + \frac{u}{1-u-x}.$$

Hence

$$C(x) < K + K \int_0^{\frac{1}{2}(1-x)} (1-x)/(1-u-x)\, du < K.$$

The integrals $A(x)$ and $B(x)$ may be treated similarly, so that $F(x)$ is bounded for x in $[0, 1)$. The same arguments may clearly be used for x in $(-1, 0]$. This completes the proof of the theorem.

CORNELL UNIVERSITY.

BIBLIOGRAPHY.

[1] S. Banach, *Théorie des opérations linéaires*, Warsaw, 1932.

[2] S. Bernstein, "Sur les polynômes orthogonaux relatifs à un segment fini," *Journal de Mathematiques*, ser. (9), vol. 9 (1930), pp. 127-177; vol. 10 (1931), pp. 219-286.

[3] I. W. Busbridge, "A theory of general transforms for functions of the class $L^p(0, \infty)$ $(1 < p \leqq 2)$," *Quarterly Journal of Mathematics* (Oxford series), vol. 9 (1938), pp. 148-160.

[4] W. B. Caton and E. Hille, "Laguerre polynomials and Laplace integrals," *Duke Mathematical Journal*, vol. 12 (1945), pp. 217-242.

[5] U. Dini, *Serie de Fourier*, Pisa, 1880, pp. 190-277.

[6] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, Cambridge, 1933.

[7] S. Kacmarz and H. Steinhaus, *Theorie der Orthogonalreihen*, Warsaw, 1935.

[8] H. Kober, "A note on approximation by rational functions," *Proceedings of the Edinburgh Mathematical Society*, vol. 7 (1946), pp. 123-133.

[9] T. M. MacRobert, "Asymptotic expressions for the Bessel functions and the Fourier-Bessel expansion," *ibid.*, vol. 39 (1921), pp. 13-20.

[10] R. E. A. C. Paley, "A remarkable system of orthogonal functions," *Proceedings of the London Mathematical Society*, vol. 34 (1932), pp. 241-279.

[11] H. Pollard, "The mean convergence of orthogonal series of polynomials," *Proceedings of the National Academy of Sciences U. S. A.*, vol. 32 (1946), pp. 8-10.

[12] ———, "The mean convergence of orthogonal series. I," *Transactions of the American Mathematical Society*, vol. 62 (1947), pp. 387-403.

[13] ———, "The mean convergence of orthogonal series. II," *ibid.*, vol. 63 (1948), pp. 355-367.

[14] ———, "The mean convergence of orthogonal series. III," *Duke Mathematical Journal*, vol. 16 (1949), pp. 189-191.

[15] M. Riesz, "Sur les fonctions conjugées," *Mathematische Zeitschrift*, vol. 27 (1927), pp. 218-244.

[16] J. Schauder, "Eine Eigenschaft der Haarschen Orthogonalsystem," *ibid.*, vol. 28 (1928), pp. 317-320.

[17] G. Szegö, *Orthogonal Polynomials*, American Mathematical Society Colloquium Publications, vol. 23, New York, 1939.

[18] E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals*, Oxford, 1937.

[19] G. N. Watson, *A Treatise on the Theory of Bessel Functions*, Cambridge, 1944.

[20] E. T. Whittaker and G. N. Watson, *A Course in Modern Analysis*, Cambridge, 1947.

[21] W. H. Young, "On series of Bessel functions," *Proceedings of the London Mathematical Society*, ser. (2), vol. 18 (1920), pp. 163-200.

ON THE STRUCTURE OF SOME SUBGROUPS OF THE MODULAR GROUP.*

By Emil Grosswald.

1. 1. Let us consider the group Γ of modular substitutions, where elements are the matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, where a, b, c, d are rational integers, satisfying

$$ad - bc = 1. \tag{1}$$

On account of the significance of a, b, c, d as coefficients of a linear fractional transformation, we consider the elements $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and $\begin{pmatrix} -a & -b \\ -c & -d \end{pmatrix}$ of Γ as identical. Imposing suitable supplementary conditions on a, b, c, d, we can define various subgroups of Γ. If, in particular, the supplementary condition is

$$c \equiv 0 \pmod{p}, \tag{2}$$

the corresponding matrices form a subgroup $\Gamma_0(p)$ of Γ, whose structure has been studied by H. Rademacher [4].[1] If, instead of (2),

$$b \equiv 0 \pmod{p} \tag{3}$$

holds, the results are perfectly analogous, as pointed out by H. Rademacher in the last paragraph of his above-mentioned paper.

In a similar way, we define the subgroups $\Gamma_0{}^0(p)$ by the conditions

$$b \equiv c \equiv 0 \pmod{p}; \tag{4}$$

and we define the "principal subgroups" $\Gamma(p)$ by the conditions

$$b \equiv c \equiv 0 \pmod{p}; \quad a \equiv d \equiv 1 \pmod{p}. \tag{4'}$$

It is our purpose to study the structure of these subgroups, indicating for any given prime number p, the corresponding set of independent generators and independent defining relations for $\Gamma_0{}^0(p)$; hereby, the abstract structure of the corresponding subgroup will be completely defined. Moreover, though

---

* Received January 17, 1950.

[1] Numbers in square brackets refer to the bibliography, at the end of the paper.

the method does not permit to write directly the independent generators as functions of p, we can find, explicitly, the generators for every particular, numerical value of p.

1.2. The main results established in this paper are:

(a) The subgroups $\Gamma_0^0(2)$, $\Gamma_0^0(3)$ are generated respectively by 2, 3 independent generators; for $p > 3$, $\Gamma_0^0(p)$ is generated by

$$m = 2[(p+2)(p-1)/12] + 3$$

independent generators. The explicit form of these generators is found.

(b) For $p = 2$, $p = 3$, $\Gamma_0^0(2)$, $\Gamma_0^0(3)$ respectively are the free groups of 2, 3 generators respectively; for $p > 3$, the generators satisfy the following defining relations:

(i) Two of the generators satisfy

$$(5) \qquad V^2(k_1', \lambda_1') = I; \qquad V^2(k_2', \lambda_2') = I \text{ if } p \equiv 1 \pmod 4.$$

(ii) Two of the generators satisfy

$$(6) \qquad V^3(k_1'', \lambda_1'') = I; \qquad V^3(k_2'', \lambda_2'') = I \text{ if } p \equiv 1 \pmod 3.$$

(iii) Two of the generators satisfy relations like (5), and two satisfy relations like (6), if $p \equiv 1 \pmod{12}$.

(iv) No defining relations exist if $p \equiv -1 \pmod{12}$; $\Gamma_0^0(p)$ is then the free group of m independent generators.

(c) For $p = 2$, both independent generators of $\Gamma_0^0(2)$ are parabolic; for $p \geqq 3$, three of the independent generators of $\Gamma_0^0(p)$ are parabolic.

(d) The genus of a fundamental region corresponding to $\Gamma_0^0(2)$ and $\Gamma_0^0(3)$ is zero; for $p > 3$, the genus is given by

$$g = [(p+2)(p-1)/12] - 1 + r(r^2 - 25)/24,$$

where $p \equiv r \pmod{12}$; $-5 \leqq r \leqq +5$.

(e)[2] If $p = 2$, the principal subgroup $\Gamma(2)$ is identical with the subgroup $\Gamma_0^0(2)$; for $p = 3$, $\Gamma(3)$ is the free group of 3 generators; for $p > 3$, and $p \equiv -1 \pmod{12}$, $\Gamma(p)$ is a free group of $p(p^2-1)/12 + 1$ generators.

The method follows closely that of Professor H. Rademacher [4].

---

[2] After the present paper had been completed, it was brought to our attention that the structure of the principal subgroups had been completely determined by H. Frasch (see [7]) who proved the results from (e) without the restriction $p \equiv -1 \pmod{12}$.

Results of K. Reidemeister [5] and O. Schreier [6] have also been used. We wish to take this opportunity to express our gratitude to Professor H. Rademacher, who suggested the subject, and whose methods and advice have made this paper possible.

2. 1. If H is a subgroup of the group G, we may write G as the sum of H and its co-sets, as

$$G = H + HR_1 + HR_2 + \cdots + HR_n + \cdots, \tag{7}$$

where

$$R_n \varepsilon G, \quad R_n \not\varepsilon H \qquad n = 1, 2, \cdots. \tag{8}$$

With $R_0 = I$, we may write (7) as

$$G = \sum_{n=0}^{N} H \cdot R_n \qquad N \leqq +\infty. \tag{7'}$$

The elements of the set $\{R_n\}$, including $R_0 = I$, are called the "representatives" of the decomposition (7′). Given any element $M \varepsilon G$, let HR_m be the co-set to which it belongs; then, following Schreier, we shall wirte $\bar{M}$ for the corresponding representative R_m. K. Reidemeister [5] has shown that if $\{S_i\}$ $(i = 1, 2, \cdots)$ are the generators of G, then all generators of H are found, eliminating from the expressions

$$R_n S_i \overline{R_n S_i}^{-1} \tag{9}$$

those, which reduce to the identity, when, in (9), the R_n and S_i range independently over all representatives $\{R_n\}$ of the decomposition (7′), respectively over all generators $\{S_i\}$ of G.

In order to be able to apply the results of O. Schreier [6] we have to take care that the system of representatives $\{R_n\}$ be chosen in such a way, as to satisfy Schreier's condition:[3] "Whenever $R_n = \prod_{i=1}^{n} S_i^{\epsilon_i}$ is a representative of its co-set, also all sections of the product, $\prod_{i=1}^{j} S_i^{\epsilon_i}$ $(j = 1, 2, \cdots, n-1)$ shall be representatives of their co-sets." Once this condition is satisfied, all defining relations of H are found (see [6] and [4]) by

$$R_n p(S_i) R_n^{-1} = I, \tag{10}$$

where the R_n range over all elements of $\{R_n\}$ (including $R_0 = I$) and

---

[3] See [6], page 177, condition "F."

$p(S_i) = I$ are the defining relations of G, and where we express the left-hand member by the generators (9) of H.

2. 2. On account of the preceding results, we shall proceed in the following order:

1) A system of representatives of the decomposition

$$\Gamma = \sum_{n=0}^{N} \Gamma_0{}^0(p) R_n \tag{11}$$

is defined, satisfying Schreier's condition.

2) By (9), the generators of $\Gamma_0{}^0(p)$ are found, using the known generators of Γ:[4]

$$S = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \text{ and } T = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{12}$$

3) By (10), the defining relations of $\Gamma_0{}^0(p)$ are determined, using the well-known fact,[4] that the defining relations of Γ are:

$$T^2 = I \quad \text{and} \quad (TS)^3 = I. \tag{13}$$

4) Use is made of the defining relations, in order to eliminate as many generators as possible, until only independent generators and independent defining relations are left.

5) The number of independent generators is then found as a function of p, and also the form and number of the defining relations established in their dependence on p.

6) As a consequence of preceding results, a relation will be found, relating directly the genus of a fundamental region, corresponding to the subgroup $\Gamma_0{}^0(p)$, to the prime number p.

3. The representatives.

Let

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \varepsilon\, \Gamma_0{}^0(p) \quad \text{and} \tag{14}$$

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = R, \qquad \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} R \,\varepsilon\, \Gamma. \tag{15}$$

Then, from

---

[4] See, e. g. [4].

$$(16)\qquad \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \begin{pmatrix} a\alpha + b\gamma & a\beta + b\delta \\ c\alpha + d\gamma & c\beta + d\delta \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix},$$

using (14) and (4), we obtain:[5]

$$(17)\qquad a\alpha \equiv A \qquad a\beta \equiv B \qquad d\gamma \equiv C \qquad d\delta \equiv D.$$

By simple, algebraic computations, using (17), we find the following results:

3. 1. If $A \equiv 0$ (implying, by (1) and (15), that $BC \equiv -1$, so that $B \not\equiv 0$, $C \not\equiv 0$), then (16) can be satisfied with $R = TS^k = \begin{pmatrix} 0 & -1 \\ 1 & k \end{pmatrix}$, where k has to satisfy

$$(18)\qquad D \equiv Ck.$$

Adding to (18) the condition

$$(19)\qquad 0 \leqq k \leqq p - 1,$$

the value of k is uniquely determined. We remark, that for $D \equiv 0$, $k \equiv 0$ and $R = T$; for $D \not\equiv 0$,

$$(19')\qquad 1 \leqq k \leqq p - 1.$$

Once k is determined by (18) and (19), or (19′), we find the corresponding matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} R^{-1} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$
$$\times S^{-k}T^{-1} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} k & 1 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} Ak - B & A \\ Ck - D & C \end{pmatrix}.$$

On account of $A \equiv 0$, we check, by (18), that actually $b = A \equiv 0$ $c = Ck - D \equiv 0$, so that indeed: $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \varepsilon\, \Gamma_0{}^0(p)$, as it had to be.

3. 2. In an entirely similar way, still using (17), we find, if $B \equiv 0$: $R = TS^kT = \begin{pmatrix} 1 & 0 \\ -k & 1 \end{pmatrix}$ with k satisfying

$$(20)\qquad kD + C \equiv 0 \qquad 1 \leqq k \leqq p - 1,$$

---

[5] Unless otherwise specified, all congruences have to be understood (mod p). For simplicity, we shall consistently suppress these words, except where confusion could arise.

and, as before, once k has been found by (20), the corresponding matrix[6] $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is found by $\begin{pmatrix} A & B \\ C & D \end{pmatrix} R^{-1} = \begin{pmatrix} A+Bk & B \\ C+Dk & D \end{pmatrix} \equiv \begin{pmatrix} A & 0 \\ 0 & D \end{pmatrix} \varepsilon \Gamma_0{}^0(p)$.

3. 3. Similarly, if $C \equiv 0$ (implying $A \not\equiv 0$, $D \not\equiv 0$), then, by (17), $R = S^l = \begin{pmatrix} 1 & l \\ 0 & 1 \end{pmatrix}$, where l satisfies

$$Al \equiv B, \qquad 0 \leqq l \leqq p-1. \tag{21}$$

If, in particular, also $B \equiv 0$, then $R = I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, since $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ then belongs to $\Gamma_0{}^0(p)$. Excluding this case, l has to satisfy

$$1 \leqq l \leqq p-1 \tag{21$'$}$$

Once l determined by (21) or (21′); the corresponding matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is found as before.

3. 4. Finally, in the general case, when $A \not\equiv 0$, $B \not\equiv 0$, $C \not\equiv 0$, we find, still using (17), that $R = TS^kTS^l = \begin{pmatrix} 1 & l \\ -k & 1-kl \end{pmatrix}$, where k and l are determined by

$$k \equiv -AC, \qquad 1 \leqq k \leqq p-1, \tag{22}$$

$$Al \equiv B, \qquad 1 \leqq l \leqq p-1. \tag{23}$$

The last relation of (17) shows also that

$$kl \equiv 1 \tag{24}$$

if and only if $D \equiv 0$.

All possible cases are now exhausted and we arrange, conveniently, the found representatives into the following array (I), which shows that they satisfy Schreier's condition:

$$\text{(I)} \qquad \begin{tikzcd}
 & S \arrow[r, dash] \arrow[dl, dash] & I \arrow[r, dash] & T \arrow[r, dash] & TS \arrow[r, dash] \arrow[d, dash] & TS^2 \arrow[r, dash] \arrow[d, dash] & \cdots \arrow[r, dash] & TS^{p-1} \arrow[d, dash] \\
S^2 \arrow[d, dash] & & & & TST \arrow[d, dash] & TS^2T \arrow[d, dash] & & TS^{p-1}T \arrow[d, dash] \\
S^3 \arrow[d, dash] & & & & TSTS \arrow[d, dash] & TS^2TS \arrow[d, dash] & & TS^{p-1}TS \arrow[d, dash] \\
S^4 & & & & TSTS^2 & TS^2TS^2 & & TS^{p-1}TS^2 \\
\vdots & & & & \vdots & \vdots & & \vdots \\
S^{p-1} & & & & TSTS^{p-1} & TS^2TS^{p-1} & & TS^{p-1}TS^{p-1}.
\end{tikzcd}$$

---

[6] By $\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \equiv \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$ we understand $a_{ij} \equiv b_{ij}$ for all i and j.

4.1. The generators of $\Gamma_0^0(p)$. From (9) we obtain, on account of (12) the generators

$$(25) \qquad RS\,\overline{RS}^{-1} \qquad\qquad (26) \qquad RT\,\overline{RT}^{-1}.$$

where R takes successively all values of the array (I). Using (25), we obtain for $R = I$.

$$IS\,\overline{IS}^{-1} = S\,\bar{S}^{-1} = S\,S^{-1} = I.$$

Similarly, as simple computations show, we obtain again the identity, if we take:

$$\begin{array}{ll}
R = S^k & 1 \leqq k \leqq p-2 \\
R = TS^k & 0 \leqq k \leqq p-2 \\
R = TS^kT & 1 \leqq k \leqq p-1 \\
R = TS^kTS^l & 1 \leqq k \leqq p-1; \quad 0 \leqq l \leqq p-2.
\end{array}$$

There remain, therefore, still to be considered, only the cases

(i) $R = S^{p-1} : RS = S^p \varepsilon \Gamma_0^0(p)$, so that $\overline{RS} = I$ and

$$(27) \qquad RS\,\overline{RS}^{-1} = S^p I^{-1} = S^p;$$

(ii) $R = TS^{p-1} : RS = TS^p = \begin{pmatrix} 0 & -1 \\ 1 & p \end{pmatrix}$, so that, by (18), (19), $\overline{RS} = T$, and

$$(28) \qquad RS\,\overline{RS}^{-1} = TS^pT^{-1};$$

(iii) $R = TS^kTS^{p-1} : RS = TS^kTS^p = \begin{pmatrix} 1 & p \\ -k & 1-kp \end{pmatrix}$, so that, by (20), $\overline{RS} = TS^{k'}T$, where k' has to satisfy $k'(1-kp) - k \equiv 0$, $1 \leqq k' \leqq p-1$, wherefrom, using the second relation from (20), $k' = k$, so that $\overline{RS} = TS^kT$ and

$$(29) \qquad RS\,\overline{RS}^{-1} = TS^kTS^pT^{-1}S^{-k}T^{-1}.$$

4.2. Let us consider now the expression (26). For

$$\begin{array}{ll}
R = I, & \\
R = TS^k, & 0 \leqq k \leqq p-1 \\
R = TS^kT, & 1 \leqq k \leqq p-1,
\end{array}$$

$RT = \overline{RT}$, so that $RT\,\overline{RT}^{-1} = I$. We have still to consider:

(i) $R = S^k : RT = S^kT = \begin{pmatrix} k & -1 \\ 1 & 0 \end{pmatrix}$, so that, by (22), (23), (24),

$\overline{RT} = TS^{k'}TS^{l'}$ with

$$k' \equiv -k, \qquad 1 \leqq k' \leqq p-1 \tag{30}$$
$$kl' \equiv -1, \qquad 1 \leqq l' \leqq p-1.$$

Let us consistently mark by k_*, the number x, satisfying the two relations (32):

$$kx \equiv -1, \qquad 1 \leqq x \leqq p-1. \tag{32}$$

Then we obtain from (30), (31):

$$l' = k_*,\ k' = p-k \text{ so that } \overline{RT} = TS^{p-k}TS^{k_*} \text{ and} \tag{33}$$

$$RT\,\overline{RT}^{-1} = S^kTS^{-k_*}T^{-1}S^{-p+k}T^{-1}. \tag{34}$$

If $R = TS^kTS^l$, then $RT = TS^kTS^lT = \begin{pmatrix} l & -1 \\ 1-kl & k \end{pmatrix}$ and two cases arise:

(ii) let $1-kl \equiv 0$; then, by **3. 3**, $\overline{RT} = S^{l'}$ with $ll' \equiv -1$, $1 \leqq l' \leqq p-1$, so that, by (32), $l' = l_*$ and $RT = S^{l_*}$. Then, as $kl \equiv 1$, $(-k)$, $l \equiv -1$, $k \equiv -l_*$, $k = p - l_*$, so that

$$RT\,\overline{RT}^{-1} = TS^{p-l_*}TS^lT \cdot S^{-l_*}. \tag{35}$$

(iii) Let $1-kl \equiv 0$. Then, by **3. 4**:

$$\overline{RT} \doteq TS^{k'}TS^{l'}, \text{ with } ll' \equiv -1,\ 1 \leqq l' \leqq p-1,$$
$$k' \equiv -l(1-kl) \equiv l(kl-1),\ 1 \leqq k' \leqq p-1.$$

Given any two numbers k and l, $1-kl \not\equiv 0$, we shall write consistently k_1 for the number x satisfying

$$x \equiv l(kl-1) \qquad 1 \leqq x \leqq p-1; \tag{36}$$

therefore, we can write using also (32),

$$l' = l_*, \quad k' = k_1, \quad \text{and} \quad \overline{RT} = TS^{k_1}TS^{l_*}, \text{ so that:}$$

$$RT\,\overline{RT}^{-1} = TS^kTS^lTS^{-l_*}T^{-1}S^{-k_1}T^{-1}. \tag{37}$$

We shall designate, conveniently, the generators of $\Gamma_0{}^0(p)$ by capital letters,[7] marking also the parameters on which they depend in each case. The results from (28)-(37) can, then, be condensed into the following array (II):

---

[7] In the choice of these letters, we have been guided by the endeavour to match, as far as possible, the designations of H. Rademacher, [4] in order to permit an easy comparison of the results.

$$\text{(II)}\qquad \begin{array}{rll} Z &= S^p & \\ U &= TS^pT^{-1} & \\ X(k) &= TS^kTS^pT^{-1}S^{-k}T^{-1} & 1 \leqq k \leqq p-1 \\ Y(k) &= S^kTS^{-k_*}T^{-1}S^{-p+k}T^{-1} & 1 \leqq k \leqq p-1 \\ W(k) &= TS^{p-k_*}TS^kTS^{-k_*} & 1 \leqq k \leqq p-1 \\ V(k,l) &= TS^kTS^lTS^{-l_*}T^{-1}S^{-k_1}T^{-1} & \begin{cases} 1 \leqq k \leqq p-1 \\ 1 \leqq l \leqq p-1 \\ kl \not\equiv 1. \end{cases} \end{array}$$

We may remark at once, that the generators $V(k,l)$ exist only for $p \geqq 3$, as for $p=2$, $k=l=1$ and $kl=1$.

5.1. The defining relations. The relations (10) become, by (13),

$$(38)\quad RT^2R^{-1} = I, \qquad \text{and} \qquad (39)\quad R(TS)^3R^{-1} = I,$$

where R ranges over all representatives of the array (I) and the left-hand sides have to be expressed as products of the generators (II). Let us consider first (38). For $R=I$ and $R=T$ we obtain, after simple computations, the trivial result $U \cdot U^{-1} = I$. Similarly, for $R = TS^k$ and $R = TS^kT$, (38) reduces to the identity $X(k) \cdot X^{-1}(k) = I$. For $R = TS^kTS^l$, however, we obtain successively:

$$\begin{aligned} TS^kTS^l \cdot T^2 \cdot S^{-l}T^{-1}S^{-k}T^{-1} = {} & TS^kTS^lTS^{-l_*}T^{-1}S^{-k_1}T^{-1} \\ & \cdot TS^{k_1}TS^{l_*}TS^{-l}T^{-1}S^{-k}T^{-1} = I. \end{aligned}$$

$$(40)\qquad V(k,l) \cdot TS^{k_1}TS^{l_*}TS^{-l}T^{-1}S^{-k}T^{-1} = I.$$

By (36), with $k = k_1$ and $l = l_*$, we obtain successively:

$$\begin{aligned} (k_1)_1 \equiv l_*(k_1l_* - 1) &\equiv l_*\{l(kl-1)l_* - 1\} \\ &\equiv l_*(-kl+1-1) \equiv -kll_* \equiv k. \end{aligned}$$

As, obviously, also $(l_*)_* = l$, it follows that

$$TS^{k_1}TS^{l_*}TS^{-l}T^{-1}S^{-k}T^{-1} = TS^{k_1}TS^{l_*}TS^{-l_{**}}T^{-1}S^{-(k_1)_1}T^{-1} = V(k_1, l_*),$$

by (II), and (40) becomes

$$(41)\qquad V(k,l) \cdot V(k_1, l_*) = I.$$

5.2. With $R = S^k$ we obtain, from (38)

$$S^kT^2S^{-k} = S^kT \cdot S^{-k_*}T^{-1}S^{-p+k}T^{-1} \cdot TS^{p-k}TS^{k_*} \cdot TS^{-k} = I, \tag{42}$$

or, by (II):

$$Y(k) \cdot W(k_*) = I, \tag{42}$$

and that is all we obtain from (38).

5.3. From (39) we obtain for

$$R = I : I(TS)^3I^{-1} = TSTSTS = TSTSTS^{-p+1} \cdot S^p = I,$$

so that, by (II)

$$W(1) \cdot Z = I. \tag{43}$$

$$\begin{aligned} R = T : T(TS)^3T^{-1} &= T^2STSTST^{-1} = TS^pT^{-1} \cdot TS^{-p}T \cdot STSTST^{-1} \\ &= U \cdot U^{-1} \cdot STSTST^{-1} = STS^{-p+1}T^{-1}S^{-p+1}T^{-1} \cdot TS^{p-1}TS^pTST^{-1} \\ &= Y(1) \cdot TS^{p-1}TS^pT^{-1}S^{-p+1}T^{-1} \cdot U = I, \end{aligned}$$

by repeated use of the definitions (II), or, finally,

$$Y(1) \cdot X(p-1) \cdot U = I; \tag{44}$$

$R = TS$ leads again to (43). $R = TS^k$, $2 \leqq k \leqq p-1$, leads to

$$\begin{aligned} &TS^k \cdot TSTSTS \cdot S^{-k}T^{-1} = TS^kTSTSTS^{-k+1}T^{-1} = TS^kTSTS^{-p+1}T^{-1}S^{-k_1}T^{-1} \\ &\cdot TS^{k_1}TS^pTS^{-k+1}T^{-1} = V(k, 1) \cdot TS^{k_1}TS^pTS^{-k+1}T^{-1} = I. \end{aligned}$$

Observing that, with $l = 1$, by (36), $k_1 = l(kl-1) = k-1$, the last equality becomes, by (II),

$$V(k, 1) \cdot X(k-1) = I \qquad 2 \leqq k \leqq p-1. \tag{45}$$

$R = TS^kT$, $1 \leqq k \leqq p-2$ leads to

$TS^kT \cdot (TS)^3 \cdot T^{-1}S^{-k}T^{-1} = TS^{k+1}TSTST^{-1}S^{-k}T^{-1} = I$, whence

$V(k+1, 1) \cdot X(k) = I$, $1 \leqq k \leqq p-2$, same as (45). Also, for $k = p-1$, $R = TS^{p-1}T$ leads to nothing new, as (39) becomes $U \cdot Y(1) \cdot X(p-1) = I$, not essentially distinct from (44).

$R = TS^kTS$ leads again to (45).

5.4. $R = TS^kTS^l$ with k and l satisfying

$$2 \leqq l \leqq p-1; \qquad 1 \leqq k \leqq p-1, \tag{46}$$

leads to the following three separate cases:

5. 4. 1. If

(47) $kl \equiv 1$ then, by (32) and (46), $l = p - k_*$ and (39) becomes

$$\begin{aligned} &TS^kTS^{p-k_*} \cdot TSTSTS \cdot S^{-p+k_*}T^{-1}S^{-k}T^{-1} \\ &\quad = TS^kTS^{p-k_*}T^{-1}S^{-p+k} \cdot S^{p-k+1}TSTS^{-p+k_*+1}T^{-1}S^{-k}T^{-1} \\ &\quad = Y^{-1}(p-k) \cdot S^{p-k+1}TS^{-(p-k+1)_*}T^{-1}S^{-k+1}T^{-1} \\ &\qquad \cdot TS^{k-1}TS^{1+(p-k+1)_*}TS^{-p+k_*+1}T^{-1}S^{-k}T^{-1} = I, \end{aligned}$$

whence, by (II),

$$Y^{-1}(p-k) \cdot Y(p-k+1) \cdot V(k-1 \cdot \bar{l}) = I,\ 2 \leqq k \leqq p-1, \tag{48}$$

with

$$\bar{l} \equiv 1 + (p-k+1)_*,\ 1 \leqq \bar{l} \leqq p-1. \tag{49}$$

In proving (48) and (49), we have used

$$\bar{l}_* \equiv \{1 + (p-k+1)_*\}_* \equiv (p - k_* - 1), \qquad \text{and} \tag{50}$$

$$(k-1)_1 = k, \tag{51}$$

whose proofs follow.

Proof of (50). Writing x for the product:

$$x = \{1 + (p-k+1)_*\}(p - k_* - 1),$$

(50) is equivalent to the congruence: $x \equiv -1$; multiplying x by $(p-k+1)$, we obtain successively:

$$\begin{aligned} x(p-k+1) &\equiv \{p-k+1-1\}(p-k_*-1) \equiv -k(-k_*-1) \\ &\equiv -1+k \equiv -(p-k+1), \end{aligned}$$

wherefrom $x \equiv -1$ follows, proving (50).

Proof of (51). By straightforward computation, and using the definition (36), we obtain successively:

$$\begin{aligned} (k-1)_1 &\equiv \{1+(p-k+1)_*\}\{(k-1)(1+(p-k+1)_*)-1\} \\ &\equiv \{1+(p-k+1)_*\}\{-(p-k+1)(1+(p-k+1)_*)-1\} \\ &\equiv \{1+(p-k+1)_*\}\{-(p-k+1)+1-1\} \\ &\equiv -(p-k+1)\{1+(p-k+1)_*\} \\ &\equiv p+k-1+1 \equiv k, \end{aligned}$$

finishing the proof of (48).

5.4.2. If

$$(52) \qquad k(l-1) \equiv 1,$$

then $kl - k \equiv 1$, or, multiplying by k_*: $-l + 1 \equiv k_*$, $l \equiv -k_* + 1$, so that[8]

$$l = p + 1 - k_*,$$

and (39) becomes

$$\begin{aligned} &TS^kTS^l \cdot TSTSTS \cdot S^{-l}T^{-1}S^{-k}T^{-1} \\ &\quad = TS^kTS^lTS^{-l_*}T^{-1}S^{-k_1}T^{-1} \cdot TS^{k_1}TS^{l_*+1}TSTS^{-l+1}T^{-1}S^{-k}T^{-1} \\ &\quad = V(k, l) \cdot TS^{k_1}TS^{l_*} \cdot STSTS \cdot S^{-l}T^{-1}S^{-k}T^{-1} = I. \end{aligned}$$

As

$$\begin{aligned} k_1 &\equiv l(kl-1) \equiv (p+1-k_*)(k(p+1-k_*)-1) \\ &\equiv (1-k_*)(k+1-1) \equiv k+1, \end{aligned}$$

and as $STSTS = T^{-1} = T$, the above relation becomes

$$V(k, l) \cdot TS^{k_1}TS^{l_*}TS^{-l}T^{-1}S^{-k}T^{-1} = V(k, l) \cdot V(k_1, l_*) = I,$$

because

$$\begin{aligned} (k_1)_1 &\equiv l_*(k_1l_* - 1) \equiv l_*\{l(kl-1)l_* - 1\} \equiv -(kl-1)l_* - l_* \\ &\equiv k + l_* - l_* \equiv k. \end{aligned}$$

The relation found is only a particular case (corresponding to $l = p + 1 - k_*$), of that already obtained under (41).

5.4.3. Let now

$$(53) \qquad kl \not\equiv 1 \quad \textit{and} \quad k(l-1) \not\equiv 1.$$

Then (39) becomes

$$TS^kTS^l \cdot TSTSTS \cdot S^{-l}T^{-1}S^{-k}T^{-1} = I,$$

whence, using the notations

$$(54) \qquad l_{j+1} \equiv (l_j)_* + 1, \qquad l = l_0, 1 \leqq l_j \leqq p-1,$$

$$(55) \qquad k_{j+1} \equiv l_j(k_jl_j - 1), \qquad k = k_0,\ 1 \leqq k_j \leqq p-1.$$

we obtain successively

---

[8] Notice that $k_* \not\equiv 1$, as $k_* = 1$ implies $k = p - 1$, whence, by (52), $l - 1 \equiv p - 1$ contradicting (46).

$$\begin{aligned} &TS^kTS^lTS^{-l_*}T^{-1}S^{-k_1}T^{-1}\cdot TS^{k_1}TS^{l_*+1}TSTS^{-l+1}T^{-1}S^{-k}T^{-1} \\ &\quad = V(k,l)\cdot TS^{k_1}TS^{l_1}TS^{-l_{1*}}T^{-1}S^{-(k_1)_1}T^{-1}\cdot TS^{(k_1)_1}TS^{l_{1*}+1}TS^{-l+1}T^{-1}S^{-k}T^{-1} \\ &\quad = V(k,l)\cdot V(k_1,l_1)\cdot TS^{k_2}TS^{l_2}TS^{-l+1}T^{-1}S^{-k}T^{-1} = I, \end{aligned}$$

or

$$V(k,l)\cdot V(k_1,l_1)\cdot V(k_2,l_2) = I. \tag{56}$$

In order to justify the last equality, we have to show that:

$$l_{2*} = l - 1, \text{ respectively } l_3 = l_{2*} + 1 = l \tag{57}$$

and

$$(k_2)_1 = k_3 = k. \tag{58}$$

Proof of (57). We have to prove that $l_2(l-1) \equiv -1$, or, by the definition of l_2,

$$(l_{1*}+1)(l-1) \equiv -1,$$

leading to $l_{1*}l - l_{1*} + l \equiv 0$, and, successively, upon multiplication by l_1:

$$-l+1+ll_1 \equiv 0; \quad l(l_1-1) \equiv -1; \quad l(l_*+1-1) \equiv -1,$$

which is true. As all steps are reversible, (57) holds.

Proof of (58). Using repeatedly the definitions (54), (55), we obtain (58) by straightforward computation:

$$\begin{aligned} k_3 &= (k_2)_1 \equiv l_2(k_2l_2-1) \equiv (1+l_{1*})\{l_1(k_1l_1-1)(1+l_{1*})-1\} \\ &\equiv (1+l_{1*})\{l_1(k_1l_1-1)-k_1l_1+1-1\} \equiv (1+l_{1*})\{k_1l_1(l_1-1)-l_1\} \\ &\equiv (1+l_{1*})(k_1l_1l_*-l_1) \equiv k_1(l_*+1)l_*-l_1-k_1l_*+1 \\ &\equiv k_1l_*l_*-l_1+1 \equiv l_*l(kl-1)l_*-(l_*+1)+1 \\ &\equiv -(kl-1)l_*-l_* \equiv k, \text{ i. e. (58).} \end{aligned}$$

From (57), (58) it results also that, starting with a pair of numbers $(k,l) = (k_0,l_0)$, satisfying the conditions (46) and (53), the sucessive, iterated pairs of numbers have period 3, i. e. $k_{n+3} = k_n$, $l_{n+3} = l_n$. Therefore, starting, e. g. in (56), with $V(k_1,l_1)$, the corresponding relation will be:

$$V(k_1,l_1)\cdot V(k_2,l_2)\cdot V(k_3,l_3) = V(k_1,l_1)\cdot V(k_2,l_2)\cdot V(k,l) = I,$$

which is precisely equivalent to (56). We shall use this remark, when we count the number of independent relations.

5. 5. We have, finally, to consider $R = S^k$, and obtain from (39):

$$\begin{aligned} R(TS)^3R^{-1} &= S^kTSTSTS\cdot S^{-k} = S^kTSTS^{-p+1}T^{-1}S^{-k_1}T^{-1}\cdot TS^{k_1}TS^pTS^{1-k} \\ &= V(k,1)\cdot TS^{k-1}TS^pTS^{-(k-1)} = V(k,1)\cdot X(k-1) = I, \end{aligned}$$

which is valid for $2 \leqq k \leqq p-1$. This contains, however, nothing new, cf. (45). For $k=1$, the relation reduces also to one already found, namely:

$$STSTST = Y(1)\cdot X(p-1)\cdot U = I, \text{ i. e. } (44).$$

5. 6. The complete list of defining relations is, therefore:

$$\text{(III)}\quad \begin{array}{ll}
W(1)\cdot Z = I & \\
Y(1)\cdot X(p-1)\cdot U = I & \\
V(k,l)\cdot X(k-1) = I & 2 \leqq k \leqq p-1 \\
Y^{-1}(p-k)\cdot Y(p-k+1)\cdot V(k-1,\bar{l}) = I & \\
 & 2 \leqq k \leqq p-1, \bar{l} = 1 + (p-k+1)_* \\
Y(k)\cdot W(k_*) = I & 1 \leqq k \leqq p-1 \\
V(k,l)\cdot V(k_1,l_*) = I & 1 \leqq k \leqq p-1, kl \not\equiv 1 \\
 & 1 \leqq l \leqq p-1 \\
V(k,l)\cdot V(k_1,l_1)\cdot V(k_2,l_2) = I & \\
 & 1 \leqq k \leqq p-1, kl \not\equiv 1 \\
 & 2 \leqq l \leqq p-1, k(l-1) \not\equiv 1.
\end{array}$$

As already pointed out,[9] the generators $V(k,l)$ do not exist for $p=2$, and (III) reduces to the three equations:

$$W(1)\cdot Z = I \qquad Y(1)\cdot X(1)\cdot U = I \qquad Y(1)\cdot W(1) = I.$$

For $p=3$, there are only two generators $V(k,l)$, namely $V(1,2)$ and $V(2,1)$, and in (III) the last relations do not exist.[10] It is easy to see that in this case there exists exactly one relation of each type, except the last (which we saw, does not exist at all) and

$$Y(k)\cdot W(k_*) = I, \text{ which gives two relations, as } k \text{ may be 1 or 2.}$$

For $p > 3$, all relations of (III) exist effectively. This is obvious for all relations, excepting, on the one hand, (48), for which case we have to prove that $(k-1)\cdot\bar{l} \not\equiv 1$ for $2 \leqq k \leqq p-1$, on the other hand, (56), for which case we have to show that $k_1 l_1 \not\equiv 1$, $l_1 \not\equiv 1$ and $k_2 l_2 \not\equiv 1$, $l_2 \not\equiv 1$, whenever k and l satisfy (46) and (53).

---

[9] See end of paragraph **4. 2.**

[10] Because, with $l=2$, and $k=1$, or $k=2$, either $kl=1$, or $k(l-1)=1$; see also § **7. 2** particularly (69), which admits a single solution and (70) which, on account of (71) admits no solution.

For the case (48), from $\bar{l} = 1 + (p - k + 1)_*$, with $2 \leqq k \leqq p - 1$, (hence $k \not\equiv 1$), we have

$$(k - 1) \cdot \bar{l} \equiv -(p - k + 1)\{1 + (p - k + 1)_*\}$$
$$\equiv -(p - k + 1) + 1 \equiv k \not\equiv 1,$$

proving the existence of (48).

For (56), assume, contrary to the assertion, that $k_1 l_1 \equiv 1$. Substituting for k_1 and l_1 their values from (54), (55), we obtain

$l(kl - 1)(l_* + 1) \equiv 1$, whence $kl(l - 1) \equiv l$ and, as $l \not\equiv 0$, $k(l - 1) \equiv 1$, contradicting (53); proving, therefore, that $k_1 l_1 \not\equiv 1$. If now, contrary to our assertion, $k_2 l_2 \equiv 1$, it follows, as before, that

$l_1(k_1 l_1 - 1)(l_{1*} + 1) \equiv 1$, whence $k_1 l_1(l_1 - 1) \equiv l_1$, so that either $l_1 \equiv 0$, or $k_1(l_1 - 1) \equiv 1$. If $l_1 \equiv 0$, then $l_* + 1 \equiv 0$, $l \equiv 1$, contrary to (46). If $k_1(l_1 - 1) \equiv 1$, then $l(kl - 1)l_* \equiv 1$, whence $kl \equiv 0$ would follow, which is impossible by (46).[11]

Obviously, also $l_1 \equiv l_* + 1 \not\equiv 1$, as $l_* \not\equiv 0$, and similarly for l_2. That finishes the proof of our assertion, that for $p > 3$, all relations (III) exist effectively.

6. Reduction to a system of independent generators and defining relations.

6.1. For $p = 2$, (II) and (III) reduce, respectively, to

$$(\mathrm{II}') \quad \begin{aligned} Z &= S^2 \\ U &= TS^2T^{-1} \\ X(1) &= TSTS^2T^{-1}S^{-1}T^{-1} \\ Y(1) &= STS^{-1}T^{-1}S^{-1}T^{-1} \\ W(1) &= TSTSTS^{-1}; \end{aligned} \qquad (\mathrm{III}') \quad \begin{aligned} W(1) \cdot Z &= I \\ Y(1) \cdot X(1) \cdot U &= I \\ Y(1) \cdot W(1) &= I. \end{aligned}$$

The three relations (III′) permit us to eliminate three of the five generators (II′), expressing them with the help of the other two, e. g.:

$$W(1) = Z^{-1}, \qquad Y(1) = Z, \qquad X(1) = Y^{-1}(1) \cdot U^{-1} = Z^{-1} \cdot U^{-1}.$$

---

[11] This conclusion would not hold, if the congruences were taken modulo n, with n composite. Difficulties of this sort make complicated an extension of the theory to the subgroups $\Gamma_0{}^0(n)$.

We remain with the two independent generators:

$$Z = S^2 = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \text{ and } U = TS^2T^{-1} = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix},$$

and no defining relations. Therefore $\Gamma_0{}^0(2)$ *is the free group of two generators.*

6.2. For $p = 3$, (II) and (III) become, respectively:

$$(\text{II}'')\quad \begin{aligned} Z &= S^3, \\ U &= TS^3T^{-1}, \\ X(1) &= TSTS^3T^{-1}S^{-1}T^{-1} \\ X(2) &= TS^2TS^3T^{-1}S^{-2}T^{-1} \\ Y(1) &= STS^{-2}T^{-1}S^{-2}T^{-1} \\ Y(2) &= S^2TS^{-1}T^{-1}S^{-1}T^{-1} \\ W(1) &= TSTSTS^{-2} \\ W(2) &= TS^2TS^2TS^{-1} \\ V(1,2) &= TSTS^2TS^{-1}T^{-1}S^{-2}T^{-1} \\ V(2,1) &= TS^2TSTS^{-2}T^{-1}S^{-1}T^{-1}, \end{aligned}$$

$$(\text{III}'')\quad \begin{aligned} W(1)\cdot Z &= I; \\ Y(1)\cdot X(2)\cdot U &= I. \\ V(2,1)\cdot X(1) &= I \\ Y^{-1}(1)\cdot Y(2)\cdot V(1,2) &= I \\ Y(1)\cdot W(2) &= I \\ Y(2)\cdot W(1) &= I \\ V(1,2)\cdot V(2,1) &= I. \end{aligned}$$

The seven relations (III′) permit us to eliminate seven of the ten generators (II″) as follows:

By the first of (III″) : $W(1) = Z^{-1}$

" " 6-th " " : $Y(2) = W^{-1}(1) = Z$

" " 4-th " " : $Y(1) = Y(2)\cdot V(1,2) = Z\cdot V(1,2)$

" " 5-th " " : $W(2) = Y^{-1}(1) = V^{-1}(1,2)\cdot Z^{-1}$

" " 2-th " " : $X(2) = Y^{-1}(1)\cdot U^{-1} = V^{-1}(1,2)\cdot Z^{-1}\cdot U^{-1}$

" " 3-th " " : $X(1) = V^{-1}(2,1) = V(1,2)$, because,

" " 7-th " " : $V(2,1) = V^{-1}(1,2)$.

In this way, all seven relations (III′) have been used, each to eliminate one generator from (II″), expressing all by the remaining three generators

$$Z = S^3 = \begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix}; \qquad U = TS^3T^{-1} = \begin{pmatrix} 1 & 0 \\ -3 & 1 \end{pmatrix};$$

$$V(1,2) = TSTS^2TS^{-1}T^{-1}S^{-2}T^{-1} = \begin{pmatrix} 4 & 3 \\ -3 & -2 \end{pmatrix},$$

and no supplementary defining relations are left. It follows that: *the subgroup* $\Gamma_0{}^0(3)$ *is the free group of three generators.*

6. 3. In the general case, with $p > 3$, we proceed in a similar way:

By (43), the 1-st of (III): $W(1) = Z^{-1}$;

by (42), with $k = p-1$, $k_* = 1$: $Y(p-1) = W^{-1}(1) = Z$;

by (48), the 4-th in (III), which we may write, conveniently as

$Y(p-k) = Y(p-k+1) \cdot V^{-1}(k-1, \bar{l})$, or, changing k into $p-k$: $Y(k) = Y(k+1)V^{-1}(p-k-1 \cdot \bar{l}')$ we eliminate,[12] successively all $Y(k)$ for $1 \leqq k \leqq p-2$.

By (42), with $1 \leqq k \leqq p-2$, we eliminate the generators $W(k_*)$, with $2 \leqq k_* \leqq p-1$.

By (45), the third in (III), written for $2 \leqq k \leqq p-1$, we eliminate the generators $X(k)$, except $X(p-1)$;

by (44), finally, the second in (III), we eliminate also $X(p-1)$.

In this way, we have used up all relations (III), except the last two, (i. e., all except (41) and (56)) and have eliminated all generators, except Z, U and the $V(k, l)$. As Z and U do not appear in (41) or (56), they will not be eliminated and enter in no defining relation.

7. 1. Our problem has been reduced to the study of the system

$$(41) \qquad V(k,l) \cdot V(k_1, l_*) = I \qquad \begin{array}{l} 1 \leqq k \leqq p-1 \\ 1 \leqq l \leqq p-1;\ kl \not\equiv 1. \end{array}$$

$$(56) \qquad V(k,l) \cdot V(k_1, l_1) \cdot V(k_2, l_2) = I \qquad \begin{array}{l} 1 \leqq k \leqq p-1;\ kl \not\equiv 1 \\ 2 \leqq l \leqq p-1;\ k(l-1) \not\equiv 1, \end{array}$$

formed by the "binary" relations (41) and the "ternary" relations (56). In (41), for each of the $p-1$ values which k may take, l can take $p-2$ distinct values, as $kl \equiv 1$ is excluded. But the $(p-1)(p-2)$ relations obtained in this way from (41) are not all distinct. In fact, by (36),

$$(k_1)_1 \equiv l_*(k_1 l_* - 1) \equiv l_*\{l(kl-1)l_* - 1\} \equiv l_*(-kl+1-1) \equiv k$$

and $(l_*)_* = l$, so that

$$V(k_1, l_*) \cdot V((k_1)_1, (l_*)_*) = I \text{ is the same as } V(k_1, l_1) \cdot V(k, l) = I.$$

Therefore there are, generally, only $\frac{1}{2}(p-1)(p-2)$ distinct relations (36). It may happen, however, that $V(k,l) = V(k_1, l_*)$, namely if, and only if

$$(59) \qquad l = l_* \qquad \text{and} \qquad (60) \qquad k = k_1.$$

---

[12] By (49) it results, that $\bar{l}' = 1 + (k+1)_*$.

(59) means that l is a solution of the congruence

$$\lambda^2 \equiv -1. \tag{61}$$

This congruence has no solution if $(-1/p) = -1$, i. e., if $p \equiv -1 \pmod 4$; or has two solutions, if $(-1/p) = +1$, i. e., if $p \equiv 1 \pmod 4$. Let us assume that $(-1/p) = +1$, so that there are two solutions, λ'_1 and λ'_2 of (61). Then, (60) becomes: $k \equiv \lambda'(k\lambda' - 1)$, or, by (61),

$$2k \equiv -\lambda'. \tag{62}$$

As $\lambda' \not\equiv 0$, (62) has, for each, λ'_1 and λ'_2, unique solutions satisfying $1 \leqq k \leqq p-1$. Therefore, if $(-1/p) = +1$, two of the binary relations (41) are of the form

$$V^2(k'_1, \lambda'_1) = I, \qquad V^2(k'_2, \lambda'_2) = I, \tag{63}$$

and $\frac{1}{2}\{(p-2)(p-1)-2\} = \frac{1}{2}(p-2)(p-1)-1$ distinct binary relations, with $V(k,l) \not\equiv V(k_1, l_*)$ are left. If $(-1/p) = -1$, no "cyclic" relations (63) exist and there are, as seen, $\frac{1}{2}(p-2)(p-1)$ distinct, non-cyclic, binary relations (41). In all cases, therefore, there are

$$\tfrac{1}{2}\{(p-2)(p-1)-(1+(-1/p))\} \tag{64}$$

distinct, binary relations, different from the (63).

7. 2. Concerning the ternary relations (56), as $2 \leqq l \leqq p-1$, we have $p-2$ choices for l; on account of the two conditions, there are, for each l, $p-3$ choices for k; there exist, therefore, $(p-2)(p-3)$ ternary relations. But, we observed already[13] that the three relations

$$V(k,l)\cdot V(k_1,l_1)\cdot V(k_2,l_2) = I, \qquad V(k_1,l_1)\cdot V(k_2,l_2)\cdot V(k_3,l_3) = I,$$
$$V(k_2,l_2)\cdot V(k_3,l_3)\cdot V(k_4,l_4) = I,$$

are not distinct; therefore, if for every pair of numbers (k,l), we have[14] $(k_1,l_1) \neq (k,l)$, then there are exactly $(p-2)(p-3)/3$ distinct ternary relations, among distinct generators. If, however, for some pair of values $(k,l) = (k_1,l_1)$ then also $(k,l) = (k_1,l_1) = (k_2,l_2)$ as, on account of (54), (55), k_2 and l_2 are found from (k_1,l_1) in the same way, as (k_1,l_1) are found from (k,l). It results that in (56) either all three generators are distinct, or else, all three are equal, and the corresponding relation from (56) becomes

$$V^3(k,l) = I. \tag{65}$$

---

[13] See remark at end of paragraph **5. 4. 3.**

[14] Here $(k_1,l_1) = (k,l)$ means $k_1 = k$ and $l_1 = l$; $(k_1,l_1) \neq (k,l)$ means that at least one of the letters is not the same in both pairs.

The equality $(k, l) = (k_1, l_1)$ means explicitly

(66) $\quad l_1 = l_* + 1 = l$ *and* $\qquad$ (67) $\quad k_1 = l(kl - 1) = k.$

Multiplying (66) by l, we obtain, successively

(68) $\qquad -1 + l \equiv l^2$ whence $4l^2 - 4l \equiv -4$, and

(69) $\qquad (2l - 1)^2 \equiv -3.$

The congruence (69) has no solution, if $(-3/p) = -1$, i. e., if $p \equiv -1 \pmod 3$ and then, the (56) reduce to $(p-2)(p-3)/3$ distinct, non-cyclical, ternary relations, each containing three different generators. If $(-3/p) = +1$, i. e., if $p \equiv 1 \pmod 3$, then (69) has two distinct solutions; let us call them λ''_1 and λ''_2. With these values, (67) becomes $\lambda''(k\lambda'' - 1) \equiv k$, whence $k(\lambda''^2 - 1) \equiv \lambda''$, or, by (68), $k(\lambda'' - 2) \equiv \lambda''$, and multiplying by $-\lambda''_*$,

(70) $\qquad k(1 + 2\lambda''_*) \equiv 1.$

For $p > 3$, $1 + 2\lambda''_* \not\equiv 0$. In fact, from $1 + 2\lambda''_* \equiv 0$ we can conclude, successively to

$$2\lambda''(2\lambda''_* + 1) \equiv 0, \quad -4 + 2\lambda'' \equiv 0, \quad 2\lambda'' - 1 \equiv 3 \text{ and } (2\lambda'' - 1)^2 \equiv 9,$$

so that, by (69),

(71) $\qquad 12 \equiv 0 \pmod p$, impossible for $p > 3$.

For $p > 3$, (70) has, therefore, exactly one solution satisfying $1 \leqq k \leqq p - 1$, for each of the two solutions λ''_1 and λ''_2 of (69). Let us call the corresponding values of k, k''_1 and k''_2 respectively. Consequently, for $p > 3$ and $(-3/p) = +1$, there are two generators, $V(k''_1, \lambda''_1)$ and $V(k''_2, \lambda''_2)$, which satisfy cyclic relations (65), of third order. The remaining $(p-2)(p-3) - 2$ relations (56) are, by a previous remark, 3 by 3 identical, so that there are, in this case $\{(p-2)(p-3) - 2\}/3$ distinct, non-cyclical, ternary relations. As, for $(-3/p) = -1$ we have $(p-2)(p-3)/3$ such relations, it results that, for every p, there are

(72) $\qquad \frac{1}{3}\{(p-2)(p-3) - (1 + (-3/p))\}$

distinct, ternary relations, different from (65).

7.3. The total number of distinct relations, which can be used for the elimination of generators $V(k, l)$ is, by (64) and (72),

$$\tfrac{1}{2}\{(p-2)(p-1) - (1 + (-1/p))\} + \tfrac{1}{3}\{(p-2)(p-3) - (1 + (-3/p))\}.$$

Each one of these relations can be used;[15] in order to eliminate one generator. As there are $(p-1)(p-2)$ different[16] generators $V(k,l)$, we are left, after using all non-cyclical relations, with

$$N=(p-1)(p-2)-\tfrac{1}{2}\{(p-2)(p-1)-(1+(-1/p))\}$$
$$-\tfrac{1}{3}\{(p-2)(p-3)-(1+(-3/p))\}$$

independent generators. A simple transformation reduces the last expression to

$$N=(p^2+\mathrm{p}-1)/6+(3(-1/p)+2(-3/p))/6. \tag{73}$$

Here, the second term is periodic, with period 12, and it is easily found that[17]

$$3(-1/p)+2(-3/p)=\begin{cases} 5 & \text{if } p\equiv 1 \pmod{12} \\ 1 & \text{"} \quad p\equiv 5 \quad \text{"} \\ -1 & \text{"} \quad p\equiv 7 \quad \text{"} \\ -5 & \text{"} \quad p\equiv 11 \quad \text{"} \end{cases}.$$

Herefrom it follows, by (73) that

$$6N=\begin{cases} p^2+p+4 & \text{if } p\equiv 1 \pmod{12} \\ p^2+p & \text{"} \quad p\equiv 5 \quad \text{"} \\ p^2+p-2 & \text{"} \quad p\equiv 7 \quad \text{"} \\ p^2+p-6 & \text{"} \quad p\equiv 11 \quad \text{"} \end{cases} \tag{74}$$

so that, in all cases

$$N=2[(p+2)(p-1)/12]+1^{18} \tag{75}$$

To this number of independent generators $V(k,l)$ of $\Gamma_0{}^0(p)$, we still have to add the generators $Z=S^p$ and $U=TS^pT^{-1}$. Therefore, the total number of independent generators of $\Gamma_0{}^0(p)$ is

$$m=N+2=2[(p+2)(p-1)/12]+3. \tag{76}$$

7.4. In the process of eliminating generators, all non-cyclic relations have been used. As defining relations there remain, therefore, only (63) and (65), which, evidently, cannot be used to eliminate generators. Recapitulating, the defining relations,

---

[15] We suppress here the proof of the statement, as it reduces to the elementary but lengthy exposition of a systematical method of elimination, patterned exactly after the corresponding methods in [4] and [7].

[16] There are $p-1$ possible choices for k, $1\leqq k\leqq p-1$; and for each k, there are $p-2$ different possible choices for l, satisfying: $1\leqq l\leqq p-1$ and $kl\not\equiv 1$.

[17] A prime $p>3$ can be only of one of the four forms considered here.

[18] Square brackets represent here, as usual, the largest integer, not exceeding the enclosed quantity.

$$(63) \qquad V^2(k'_1, \lambda'_1) = I, \qquad V^2(k'_2, \lambda'_2) = I$$

exist, if $p \equiv 1 \pmod 4$, $\lambda'_1, \lambda'_2, k'_1, k'_2$ being the solutions of (61) and 62), respectively.

$$(65) \qquad V^3(k''_1, \lambda''_1) = I, \qquad V^3(k''_2, \lambda''_2) = I$$

exist, if $p \equiv 1 \pmod 3$, $\lambda''_1, \lambda''_2, k''_1, k''_2$ being the solutions of (69) and (70), respectively.[19] If $p \equiv 1 \pmod{12}$, then both (63) and (65) exist; if $p \equiv -1 \pmod{12}$, then there are no defining relations at all and $\Gamma_0{}^0(p)$ is the free group of m generators. The number of defining relations is, therefore, for every p,

$$(77) \qquad \nu = \{1 + (-1/p)\} + \{1 + (-3/p)\} = 2 + (-1/p) + (-3/p),$$

as results immediately, either from **7. 3**, or as a consequence of (64) and (72).

7. 5. The explicit form of the independent generators of $\Gamma_0{}^0(p)$ is easily determined, using the definitions (II), and we find

$$(78) \qquad Z = \begin{pmatrix} 1 & p \\ 0 & 1 \end{pmatrix}, \qquad U = \begin{pmatrix} 1 & 0 \\ -p & 1 \end{pmatrix},$$
$$V(k, l) = \begin{pmatrix} k_1(1 + ll_*) - l & 1 + ll_* \\ k_1 l_* + kl - 1 - kk_1(1 + ll_*) & -k(1 + ll_*) + l_* \end{pmatrix}.$$

8. 1. As is well-known[20] $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is parabolic if, and only if $a + d = \pm 2$. We see, therefore, from (78) that Z and U are parabolic for every p. As for $V(k, l)$, it is parabolic if, and only if:

$$(79) \qquad k_1(1 + ll_*) - l - k(1 + ll_*) + l_* = \pm 2 \text{ or}$$
$$(k_1 - k)(1 + ll_*) + (l_* - l) = \pm 2.$$

As $1 + ll_* = n \cdot p$, $n = 1, 2, \cdots$ and $|l_* - l| < p$, (79) can hold only if either:

(i) $k_1 - k = 0$ and $|l_* - l| = 2$

---

[19] By elementary computations it is easy to show also that:

$$\lambda'_1 + \lambda'_2 = k'_1 + k'_2 = p, \qquad \lambda''_1 + \lambda''_2 = p + 1 \text{ and } k''_1 + k''_2 = p,$$

and this remark simplifies numerical computations.

[20] See, e. g., [1], page 21.

or if:

(ii) $| k_1 - k | = 1$ and $ll_* + 1 = p$, whence, by (79), $| l_* - l | = p - 2$.
The first alternatve is impossible. In fact, substituting k_1 by its value and multiplying by $l_*, k_1 - k = 0$ becomes:

$$k(l_* + l) \equiv 1; \text{ also, from } | l_* - l | = 2 \text{ results:}^{21} \tag{80}$$

$$k(l_* - l) \equiv \pm 2k. \tag{81}$$

Adding, we obtain: $2kl_* = 1 \pm 2k$, whence, after simple transformations:

$$2k(l \pm 1) \equiv \mp l. \tag{82}$$

Subtracting (81) from (80) we obtain: $2kl \equiv 1 \mp 2k$, so that:

$$2k(l \pm 1) \equiv 1. \tag{83}$$

Comparing (82) with (83), it results that $l = 1$, or $l = p - 1$, leading, respectively, to $l_* = p - 1$, or $l_* = 1$, so that $| l_* - l | = p - 2 \neq 2$, proving that the first alternative is impossible.

In the second alternative, the system $ll_* = p - 1$, $l_* - l = \pm (p - 2)$ is readily solved and has the solutions

$$l = 1, \quad l_* = p - 1 \text{ and } l = p - 1, \quad l_* = 1.$$

As both solutions actually lead to $k_1 = k - 1$, $k_1 = k + 1$, respectively, the corresponding $V(k, 1)$ and $V(k, p - 1)$ are, effectively, parabolic.

8.2. By (41), $V(k, 1) = V^{-1}(k_1, p - 1)$ and no parabolic generators of the form $V(k, 1)$ need appear in the set of independent generators of $\Gamma_0{}^0(p)$. Of the $p - 2$ parabolic generators. $V(k_1, p - 1)$ $(k_1 = 1, 2, \cdots, p - 2)$, we can eliminate all but one, using the relations (56). For that, let us list them, starting with $l = 2$; to it corresponds $l_* = (p - 1)/2$ and $l_1 = l_* + 1 = (p + 1)/2$; $l_{1*} = p - 2$ and $l_2 = l_{1*} + 1 = p - 1$. Therefore, for $V(k, 2)$, taken as first factor in (56), the last factor is $V(k_2, p - 1)$, where k_2 takes all integral values $1 \leqq k \leqq p - 2$, except $(p - 1)/2$, as $(p - 1)/2 \cdot (p - 2) = p(p - 3)/2 + 1 \equiv 1$, contradicting (53). From preceding relations it is also easy to see, that in any ternary relations (56), there can be at most one parabolic factor; consequently, we can use the $p - 3$ relations, starting with $V(k, 2)$, in order to eliminate $p - 3$ of the $p - 2$ parabolic generators $V(k, p - 1)$ and are left only with $V(\frac{1}{2}(p - 1), p - 1)$, which cannot be eliminated. Consequently, it follows that among the independent generators

---

[21] Here, whenever double signs are written, we have to consider either all upper, or all lower signs.

of every subgroup $\Gamma_0{}^0(p)$, $p > 3$, there are exactly three parabolic generators, namely:

$$Z = S^p = \begin{pmatrix} 1 & p \\ 0 & 1 \end{pmatrix}, \qquad U = TS^pT^{-1} = \begin{pmatrix} 1 & 0 \\ -p & 1 \end{pmatrix} \text{ and}$$

$$V(\tfrac{1}{2}(p-1), p-1) = \begin{pmatrix} \frac{1}{2}p(p-1)+1 & p \\ -\frac{1}{4}p(p-1)^2 & -\frac{1}{2}p(p-1)+1 \end{pmatrix}.$$

9. Using the relations (61), (62), (63), (65), (69), (70) and (76), the following results are found for the subgroups $\Gamma_0{}^0(p)$, corresponding to the first few values of p:

	Generators	*Defining Relations*
$p = 2$	Z, U	none
$p = 3$	$Z, U, V(1, 2)$	none
$p = 5$	$Z, U, V(2, 4)$; $V(4, 2), V(1, 3)$ $V(2, 2), V(4, 3)$	$V^2(4, 2) = I$, $V^2(1, 3) = I$.
$p = 7$	Z, U, other 9 indep. gen. $V(k, l)$	$V^3(3, 3) = I$, $V^3(4, 5) = I$.
$p = 11$	Z, U, " 21 " " $V(k, l)$	none
$p = 13$	Z, U, " 31 " " $V(k, l)$	$V^2(4, 5) = I$, $V^2(9, 8) = I$, $V^3(11, 10) = I$, $V^3(2, 4) = I$.
$p = 17$	Z, U, " 57 " " $V(k, l)$	$V^2(15, 4) = I$, $V^2(2, 13) = I$.
$p = 19$	Z, U, " 63 " " $V(k, l)$	$V^3(5, 12) = I$, $V^3(14, 8) = I$.
$p = 23$	Z, U, " 91 " " $V(k, l)$	none.

10. The genus of the fundamental region. Let us consider, for a given subgroup $\Gamma_0{}^0(p)$, a fundamental region[22] such, that the independent generators found in paragraphs 6-10 should establish the correspondence between pairs of its sides.

Let $2n$ be the number of sides; k the number of cycles, corresponding to fixed point (that is, exclusive of cycles of accidental corners); m and ν the number of independent generators and defining relations of $\Gamma_0{}^0(p)$ respectively; finally g the genus of the fundamental region, as defined. Then[23] $2n = 4g + 2k$, hence

$$g = \tfrac{1}{2}(n - k). \tag{84}$$

---

[22] See, e. g., [2], vol. 2, p. 4.

[23] See [2], vol. 2, p. 303 and vol. 1, p. 319. (The notations there are different from ours.)

According to the definition of the fundamental region,

$$n = m. \tag{85}$$

The number k, of cycles, equals the number of defining relations ν, to which we have to add the number of parabolic generators; as shown in **8.2** for every $p > 3$, there are exactly three independent parabolic generators. Therefore:

$$k = \nu + 3. \tag{86}$$

The values of m and ν are known from (76) and (77). Substituting these values in (85), (86) and then in (84), we obtain, for $p > 3$, the following expression for g:

$$\begin{aligned} g &= [(p+2)(p-1)/12] - \{1 + \tfrac{1}{2}((-1/p) + (-3/p))\} \\ &= [(p+2)(p-1)/12] - K(p). \end{aligned} \tag{87}$$

Here $K(n)$ is a periodic function, of period 12, such that $K(n) = 0, 1, 2,$ according as $n \equiv -1, \pm 5, +1 \pmod{12}$, respectively. As every prime number $p > 3$ satisfies one of the four congruences $p \equiv r \pmod{12}$, $-6 \leqq r \leqq 6$, with $r = \pm 1, \pm 5$ it results that $K(p) = 1 - r(r^2 - 25)/24$; so that

$$g = [(p+2)(p-1)/12] - 1 + r(r^2 - 25)/24. \tag{87'}$$

For $p = 2$ and $p = 3$ we find directly, by (84), (85) and (86) with the known values for m and ν, that $g = 0$ in both cases. (87′) is valid for $p > 3$ and using it, we find as genus of the fundamental regions, corresponding to the subgroups listed in paragraph **9**:

$$\begin{array}{llllllllll} p\colon & 2, & 3, & 5, & 7, & 11, & 13, & 17, & 19, & 23, \cdots; \\ g\colon & 0, & 0, & 1, & 2, & 10, & 13, & 24, & 30, & 45, \cdots . \end{array}$$

11. We can use the results obtained so far for subgroups $\Gamma_0{}^0(p)$, in order to establish some properties of the "principal subgroups"[24] $\Gamma(p)$ defined by (4), (4′). Evidently the $\Gamma(p)$ are subgroups of the $\Gamma_0{}^0(p)$. We may, therefore, in order to determine their structure, apply the same method used in paragraphs **3** to **8**, starting with the $\Gamma_0{}^0(p)$ as known groups, given by their generators and defining relations, as listed in paragraph **9**. In perfect analogy with (14), (15), (16), etc., we write now:

---

[24] See [7].

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \varepsilon\, \Gamma(p) \qquad \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = R\, \varepsilon\, \Gamma_0{}^0(p)$$

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} R\, \varepsilon\, \Gamma_0{}^0(p), \tag{88}$$

and have, on account of (4) and (4′):

$$b \equiv c \equiv \beta \equiv \gamma \equiv B \equiv C \equiv 0 \qquad a \equiv d \equiv 1. \tag{89}$$

Writing explicitly (88) we obtain:

$$\begin{aligned} a\alpha + b\gamma &\equiv A \\ a\beta + b\delta &\equiv B \\ c\alpha + d\gamma &\equiv C \\ c\beta + d\delta &\equiv D. \end{aligned} \tag{90}$$

On account of (1) and (89), the first of (90) reduces to

$$\alpha \equiv A; \tag{91}$$

the others reduce to trivial identities.

Let[25] $A \equiv k$, $1 \leqq k \leqq \frac{1}{2}(p-1)$. For $k=1$; (88) is satisfied by $R_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I$, and, in general, for $2 \leqq k \leqq \frac{1}{2}(p-1)$, by: $R_k = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ with $\alpha \equiv k$, $\beta \equiv \gamma \equiv 0$, $\delta \equiv p - k_*$. With these notations, we obtain the relation, similar to (7′):

$$\Gamma_0{}^0(p) = \sum_{n=1}^{(p-1)/2} \Gamma(p) \cdot R_k, \tag{92}$$

with $R_1 = I$.

For $p=2$ or $p=3$, (92) becomes, simply:

$$\Gamma_0{}^0(2) = \Gamma(2), \qquad \Gamma_0{}^0(3) = \Gamma(3), \text{ respectively.}$$

For $p > 3$ and $p \equiv -1 \pmod{12}$ we have seen in **7.4** that $\Gamma_0{}^0(p)$ is the free group of m generators, with m determined by (76). As $\Gamma(p)$ is a subgroup of $\Gamma_0{}^0(p)$, it results, by a theorem of J. Nielsen [3] and O. Schreier [6] that, choosing conveniently its generators, $\Gamma(p)$ is also a free group. The number N, of the corresponding independent generators, can be found, using a relation due also to O. Schreier [6], namely (in slightly modified notations, to correspond with ours):

---

[25] As $AD \equiv 1$, the case $A \equiv 0$ can be discarded from the start. If $\frac{1}{2}(p-1) < k \leqq p-1$, then consider $k-p$ and use $\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} -a & -b \\ -c & -d \end{pmatrix}$.

$$(93) \qquad N = j(m-1)+1,$$

where m is the number of generators of $\Gamma_0{}^0(p)$, as given by (76) and j is the index of $\Gamma(p)$ in $\Gamma_0{}^0(p)$. By (92), $j = \frac{1}{2}(p-1)$; substituting this value and that of m in (93), taking also into account that $p \equiv -1 \pmod{12}$, we obtain:

$$N = p(p^2-1)/12 + 1,$$

a value found by H. Frasch without the restriction $p \equiv -1 \pmod{12}$.

UNIVERSITY OF PENNSYLVANIA.

BIBLIOGRAPHY.

[1] L. R. Ford, *Automorphic Functions,* New York, 1929.

[2] R. Fricke und F. Klein, *Vorlesungen über die Theorie der automorphen Funktionen,* Leipzig, 1897.

[3] J. Nielsen, "On Regning med ikke-kommutative Faktorer og dens Anvendelse i Gruppeteorien," *Matematisk Tidskrift,* Afd. B (1921), pp. 77-94.

[4] H. Rademacher, "Ueber die Erzeugenden von Kongruenzuntergruppen der Modulgruppe," *Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität,* vol. 7 (1929), pp. 134-148.

[5] K. Reidemeister, "Knoten und Gruppen," *ibid.,* vol. 5 (1926), pp. 7-23.

[6] O. Schreier, "Die Untergruppen der freien Gruppen," *ibid.,* vol. 5 (1926), pp. 161-183.

[7] H. Frasch, "Die Erzeugenden der Hauptkongruenzgruppen für Primzahlstufen," *Mathematische Annalen,* vol. 108 (1933), pp. 229-252.

ON THE SOLUTION OF SUM-EQUATIONS.*

By I. M. Sheffer.

1. Introduction.

The system of linear equations

$$(1.1)\qquad b_{n0}x_0 + b_{n1}x_1 + \cdots + b_{nj}x_j + \cdots = c_n, \qquad (n = 0, 1, \cdots),$$

can, in many interesting cases, be reduced to a system of *sum-equations*[1]

$$(1.2)\qquad \mathfrak{A}: a_{n0}x_n + a_{n1}x_{n+1} + \cdots + a_{nj}x_{n+j} + \cdots = c_n \quad (n = 0, 1, \cdots).$$

Letting X denote the sequence $\{x_n\}$, the left side of (1.2) will be abbreviated $A_n[X]$. It is our purpose to study the system (1.2), in which all quantities are complex numbers.[2]

Define the functions $\{A_n(t)\}$ by

$$(1.3)\qquad A_n(t) = \sum_{j=0}^{\infty} a_{nj}t^j \qquad (n = 0, 1, \cdots).$$

A sequence $\{H_n(1/t)\}$ of polynomials in $1/t$ is introduced as biorthogonal to the set $\{t^nA_n(t)\}$, and this defines the generating function

$$(1.4)\qquad \Phi(u, 1/t) = \sum_{n=0}^{\infty} H_n(1/t)u^n.$$

The function Φ plays a central rôle in the theory. Use is also made of the functions

$$(1.5)\qquad \Psi(1/u, t) = \sum_{n=0}^{\infty} t^nA_n(t)/u^n$$

and

$$(1.6)\qquad \Lambda(u, 1/t) = \sum_{n=0}^{\infty} H_n(1/t)u^nA_n(u).$$

The results of § 2 are related to work of von Koch [9]. They state conditions under which (1.2) has a unique solution, and set the background for § 3 where it is shown how from knowledge of the analytic continuation

---

* Received January 7, 1950; revised May 2, 1950. Presented to the American Mathematical Society, October, 1949.

[1] For example, if all principal minors in (1.1) are different from zero, then (1.2) can be obtained by forming linear combinations of the equations of (1.1) in an obvious manner.

[2] Many articles have been written on sum-equations, including all the references (save that of Boas) cited in the bibliography at the end of the present work.

of the function $\Phi(u, 1/t)$ (in the variable u) one obtains solutions of the homogeneous system (1. 2) $[c_n \equiv 0]$.

2. Preliminary results. We introduce some definitions.

Definition. System $\mathcal{A}$ has the property $P_1[R]$ if $a_{n0} \neq 0$ for all n in (1. 3) and if each series of (1. 3) has a convergence radius at least as great as R $(0 < R \leqq \infty)$.

Definition. A series $\Sigma f_{j_1,\ldots,j_k}(u_1, \cdots, u_p)$ is said to *converge "uniformly"* for u_s on a point set $\mathcal{J}_s$ $(s = 1, \cdots, p)$ if it converges uniformly (in the usual sense) for u_s on $\mathcal{J}^*_s$ where $\mathcal{J}^*_s$ is an arbitrary closed point set contained in $\mathcal{J}_s$.

LEMMA 2. 1. *Let condition $P_1[R]$ hold. For each $n = 0, 1, \cdots$ there is a unique polynomial $H_n(1/t)$, of degree exactly n in $1/t$, that is orthogonal*[3] *to the set $\{t^n A_n(t)\}$ in the sense*

$$\frac{1}{2\pi i}\int_C H_n(1/t)\, t^k A_k(t)\, dt/t = \delta_{nk}. \tag{2.1}$$

Here C is any rectifiable simple closed curve[4] *surrounding $t = 0$ and lying in $| t | < R$. Moreover,*

$$a_{0n}H_0(1/t) + a_{1,n-1}H_1(1/t) + \cdots + a_{n0}H_n(1/t) = 1/t^n. \tag{2.2}$$

The proof of (2. 1) is direct, and (2. 2) is a consequence of (2. 1).

Remark. Let n be fixed. If $H_n(1/t) = \sum_{j=0}^{n} h_{nj}/t^j$ then we see from (2. 1) and (2. 6) that $\{x_j\}$ defined by $x_j = h_{nj}$, $j = 0, 1, \cdots, n$; $x_j = 0$, $j > n$ is a solution of system (1. 2) with $c_j = 0$ if $j \neq n$, $c_n = 1$. It is thus seen that the results of § 2 hold at least formally, especially those results involving series (2. 7). It would be possible to express these results (of § 2) in terms of the matrix

$$\mathcal{A} = \begin{pmatrix} a_{00} & a_{01} & a_{02} \cdots \\ 0 & a_{10} & a_{11} \cdots \\ \cdot & \cdot & \cdot \quad \cdot \\ \cdot & \cdot & \cdot \quad \cdot \end{pmatrix}; \text{ e. g., if } \mathcal{H} = \begin{pmatrix} h_{00} & \cdots & \cdot \\ h_{10} & h_{11} \cdots & \\ \cdot & \cdot \quad \cdot & \cdot \\ \cdot & \cdot \quad \cdot & \cdot \end{pmatrix},$$

then (2. 2) expresses the relation $\mathcal{H}'\mathcal{A} = \mathcal{I}$, where $\mathcal{I}$ is the identity and $\mathcal{H}'$ is the transpose of $\mathcal{H}$. For the purpose of preparing the way for § 3, however, it is desirable to use the functions Φ, Ψ.

---

[3] Boas [1] obtained a similar orthogonality relation, with a different purpose in view.

[4] Here and elsewhere, contours are traversed in the positive direction.

Definition. System $\mathfrak{A}$ has the property $P_s[R, q, r]$, $s = 2, 3, 4$ if $P_1[R]$ holds and if q, r exist so that respectively: ($s = 2$) $0 < q$, $0 \leqq r$ and the Φ-series (1.4) converges "uniformly" for $|u| < q$, $|t| > r$; ($s = 3$) $0 \leqq q$, $0 < r \leqq R$ and the Ψ-series (1.5) converges "uniformly" for $|u| > q$, $|t| < r$; ($s = 4$) $0 < q \leqq R$, $0 \leqq r$ and the Λ-series (1.6) converges "uniformly" for $|u| < q$, $|t| > r$.

Remark. It is undesirable for our purpose if the Φ-series has either a zero or infinite radius of convergence for a given t. In such case we can choose a sequence $\{\lambda_n\}$ of non-zero constants so that on replacing $\{H_n(1/t)\}$ by $\{\lambda_n H(1/t)\}$ the new Φ-series has a finite, non-zero convergence radius. This alteration replaces $\{A_n(t)\}$ by $\{A_n(t)/\lambda_n\}$, as we see from (2.1); so system (1.2) is replaced by an equivalent system, obtained by multiplying the n-th equation ($n = 0, 1, \cdots$) by $1/\lambda_n$.

LEMMA 2.2. *If $\mathfrak{A}$ satisfies $P_4[R, q, r]$ then $q \leqq r$ and*

$$\Lambda(u, 1/t) = t/(t - u). \tag{2.3}$$

Proof. For $|u| < q$ series (1.6) can be expanded in a power series in u for each t in $|t| > r$; and by (2.2) this series is $\sum_0^\infty (u/t)^n$, convergent for $|u| < q$, $|t| > r$.

LEMMA 2.3. *Let $\mathfrak{A}$ satisfy $P_2[R, q_2, r_2]$ and $P_3[R, q_3, r_3]$, with $q_3 < q_2$. Then $P_4[R, q, r]$ holds with $q = r_3$, $r = r_2$; and "uniformly" in $|u| < r_3$, $|t| > r_2$,*

$$\frac{1}{2\pi i}\int_\Gamma \Phi(s, 1/t)\Psi(1/s, u)\,ds/s = \sum_{n=0}^\infty H_n(1/t)u^n A_n(u) = t/(t - u), \tag{2.4}$$

where Γ is a rectifiable simple closed curve surrounding $s = 0$ and lying in the ring $q_3 < |s| < q_2$. Moreover, $r_3 \leqq r_2$.

Proof. Substituting (1.4), (1.5) into (2.4), the left side reduces to (1.6), "uniformly" convergent for $|u| < r_3$, $|t| > r_2$. Lemma 2.2 then applies.

Definition. By the *type* of a sequence $\{x_n\}$ is meant the number

$$((x_n)) = \limsup |x_n|^{1/n}. \tag{2.5}$$

A sequence $X = \{x_n\}$ and the series (or function) $X(1/t) = \sum_0^\infty x_n/t^n$ are uniquely determined each by the other, and when both appear it will be understood that they bear this relation to each other. It is readily seen that if system $\mathfrak{A}$ satisfies $P_1[R]$ and $\{x_n\}$ is of type $\sigma < R$, then

$$A_n[X] \equiv \sum_{j=0}^{\infty} a_{nj} x_{n+j} = \frac{1}{2\pi i}\int_C X(1/t)\, t^{n-1} A_n(t)\, dt, \tag{2.6}$$

where C *is* a rectifiable simple closed curve surrounding $t = 0$ and lying in the ring $\sigma < | t | < R$.

THEOREM 2.1. *Let $\mathfrak{A}$ satisfy $P_4[R, q, r]$. If $\{x_n\}$ is a solution of system* (1.2), *of type less than q, then*

$$X(1/t) = \sum_{n=0}^{\infty} c_n H_n(1/t), \tag{2.7}$$

the series being "uniformly" convergent in $| t | > r$.

Proof. As $((x_n)) = \sigma < q \leqq R$, series (1.6) and $X(1/u) = \sum_{0}^{\infty} x_n/u^n$ converge uniformly for $| u | = \lambda$ $(\sigma < \lambda < q)$ and $| t | \geqq r' > r$. Hence if C is the circle $| u | = \lambda$,

$$\frac{1}{2\pi i}\int_C \Lambda(u, 1/t) X(1/u)\, du/u = \sum_{n=0}^{\infty} H_n(1/t) \cdot \frac{1}{2\pi i}\int_C X(1/u)\, w^{n-1} A_n(u)\, du,$$

"uniformly" convergent for $| t | > r$. From (2.3) the left side has the value $X(1/t)$, and from (2.6) the coefficient of $H_n(1/t)$ is $A_n[X]$, which is to say, c_n.

Remark. Under the conditions of Theorem 2.1 the homogeneous system (1.2) $[c_n \equiv 0]$ has no solution of type less than q other than $x_n \equiv 0$.

THEOREM 2.2. *If $\mathfrak{A}$ satisfies $P_1[R]$ and $\{c_n\}$ is such that series* (2.7) *converges "uniformly" in $| t | > \delta$ $(\delta < R)$ then $\{x_n\}$ is a solution of system* (1.2), *and $((x_n)) \leqq \delta$.*

Proof. Multiply (2.7) by $t^{k-1}A_k(t)/(2\pi i)$ and integrate over the circle C: $| t | = \lambda$, $\delta < \lambda < R$. Using (2.1) and (2.6) we obtain $A_n[X] = c_n$; and clearly, $((x_n)) \leqq \delta$.

Remark. By using the Cauchy integral for the coefficients $\{x_n\}$ in $X(1/t)$, it is seen from (2.7) that the solution in Theorem 2.2 is given by

$$x_n = \sum_{k=n}^{\infty} \lambda_{nk} c_k \tag{2.8}$$

where

$$\lambda_{nk} = \frac{1}{2\pi i}\int_\Gamma t^{n-1} H_k(1/t)\, dt, \tag{2.9}$$

Γ being rectifiable simple closed curve surrounding $t = 0$.

THEOREM 2. 3. *Let* $\mathfrak{A}$ *satisfy* $P_2[R, q, r]$, *with* $r < R$. *For every* $\{c_n\}$ *with* $((c_n)) < \min\{q, R\}$, *system* (1. 2) *has the solution* $\{x_n\}$ *given by*

$$X(1/t) = \frac{1}{2\pi i}\int_\gamma C(1/u)\Phi(u, 1/t)\,du/u, \tag{2.10}$$

where

$$C(1/u) = \sum_{n=0}^{\infty} c_n/u^n \tag{2.11}$$

and γ *is a rectifiable simple closed curve surrounding* $u = 0$ *and lying in the ring* $((c_n)) < |u| < \min\{q, R\}$. *Moreover,* $((x_n)) \leqq r$.

Proof. The Φ series (1. 4) converges uniformly for u on γ and $|t| \geqq r' > r$, so (2. 10) becomes $X(1/t) = \sum_0^\infty c_n H_n(1/t)$, "uniformly" convergent for $|t| > r$. Theorem 2. 2 now applies.

Remark. If $\mathfrak{A}$ satisfies $P_2[R, q, r]$ then $\Phi(u, 1/t)$ can be expanded in a power series in $1/t$:

$$\Phi(u, 1/t) = \sum_{n=0}^{\infty} u^n L_n(u)/t^n, \tag{2.12}$$

"uniformly" convergent in $|u| < q$, $|t| > r$, where the functions $L_n(u)$ are analytic in $|u| < q$. In fact,

$$u^n L_n(u) = \frac{1}{2\pi i}\int_C \Phi(u, 1/t)t^{n-1}dt \tag{2.13}$$

where C is a rectifiable simple closed curve surrounding $t = 0$ and lying in $|t| > r$. The solution $\{x_n\}$ of Theorem 2. 3 can therefore be expressed in the form

$$x_n = \frac{1}{2\pi i}\int_\gamma C(1/u)u^{n-1}L_n(u)\,du. \tag{2.14}$$

3. Solutions of the homogeneous system. The results of § 2 concern system (1. 2) in a range where there is no solution of the homogeneous case $[c_n \equiv 0]$ other than $x_n \equiv 0$. In the present section it will be shown that non-zero solutions of the homogeneous system arise when the function $\Phi(u, 1/t)$, on being continued analytically in the variable u, has the representation (3. 1) or (3. 31). Thus a relation is established between the analytic properties of the function Φ on the one hand and solutions of system (1. 2) on the other.

Definition. The set of functions $\Theta_j(u)$, $j = 1, \cdots, k$ is said to be "*linearly dependent*" in $|u| < q$ if constants $a_1, \cdots, a_k$ exist, not all zero, such that the function $\sum_{j=1}^{k} a_j\Theta_j(u)$ is analytic in a circle, center $u = 0$ and radius exceeding q. In the contrary case the functions are "linearly independent" in $|u| < q$.

Definition. Let q, q_1, q' exist, with $q < q_1 < q'$, such that $\mathfrak{a}$ has the property $P_2[R, q, r]$ but not $P_2[R, q_1, r]$, and let $\Phi(u, 1/t)$ on being continued analytically have the representation

$$(3.1) \qquad \Phi(u, 1/t) = \sum_{j=1}^{k} R_j(1/t)\Theta_j(u) + \Phi^*(u, 1/t)$$

where the series

$$(3.2) \qquad \Phi^*(u, 1/t) = \sum_{n=0}^{\infty} H^*_n(1/t)u^n$$

is "uniformly" convergent in $|u| < q'$, $|t| > r$, and the functions $H^*_n(1/t)$ are analytic in $|t| > r$. Moreover, let the functions $R_j(1/t)$ be analytic and linearly independent in $|t| > r$, and the functions $\Theta_j(u)$ analytic in $|u| < q$ and "linearly independent" in $|u| < q_1$. Then the expansion (3.1) is said to be *canonical, relative to* $[q, q_1, q'; r]$.

A canonical form is esentially unique, as the following shows.

Lemma 3.1. *Let* $\Phi(u, 1/t)$ *have the two canonical forms*

$$(3.3) \qquad \Phi(u, 1/t) = \sum_{j=1}^{k(i)} R_j^{(i)}(1/t)\Theta_j^{(i)}(u) + \Phi^{*(i)}(u, 1/t) \qquad (i = 1, 2)$$

relative to $[q, q_1, q'; r]$. *Then* (i) $k(1) = k(2)$, *and* (ii) *the members of each of the two sets* $\{R_j^{(i)}\}$ $(i = 1, 2)$ *are linear combinations of the members of the other set.*

Proof. By subtraction we have from (3.3):

$$(3.4) \qquad \Phi^*(u, 1/t) = \sum_{j=1}^{k(1)} R_j^{(1)}\Theta_j^{(1)} - \sum_{j=1}^{k(2)} R_j^{(2)}\Theta_j^{(2)},$$

where $\Phi^* = \Phi^{*(2)} - \Phi^{*(1)}$. Let $k(1) \geqq k(2)$. Since set $\{R_j^{(1)}\}$ is linearly independent, there exist $k(1)$ values $t = t_1, \cdots, t_{k(1)}$ in $|t| > r$ for which the determinant $|R_j^{(1)}(1/t_s)|$ $(j, s = 1, \cdots, k(1))$ is not zero, so on taking $t = t_1, \cdots, t_{k(1)}$ in (3.4) we can solve for $\Theta_j^{(1)}(u)$:

$$(3.5) \qquad \Theta_j^{(1)}(u) = \theta_j^{(1)}(u) + \sum_{q=1}^{k(2)} \lambda_{j,q}\Theta_q^{(2)}(u) \qquad (j = 1, \cdots, k(1)),$$

where $\theta_j^{(1)}(u)$ is analytic in $|u| < q'$.

If $k(2) < k(1)$ then the matrix (λ_{jq}) is of rank less than $k(1)$, so the system of equations

$$\sum_{j=1}^{k(1)} a_j \lambda_{jq} = 0 \qquad (q = 1, \cdots, k(2)) \tag{3.6}$$

has a solution $\{a_j\}$ not identically zero. For such $\{a_j\}$ we find from (3.5) that

$$\sum_{j=1}^{k(1)} a_j \Theta_j^{(1)}(u) = \sum_{j=1}^{k(1)} a_j \theta_j^{(1)}(u). \tag{3.7}$$

But the right side is analytic in $|u| < q'$, so the set $\{\Theta_j^{(1)}(u)\}$ is "linearly dependent" in $|w| < q_1$, contrary to hypothesis. Thus, $k(1) = k(2), = k$.

Now substitute (3.5) into (3.4), absorbing the terms containing $\theta_j^{(1)}(u)$ into Φ^*:

$$\Phi^*(u, 1/t) = \sum_{q=1}^{k} \{ \sum_{j=1}^{k} \lambda_{jq} R_j^{(1)}(1/t) - R_q^{(2)}(1/t) \} \Theta_q^{(2)}(u). \tag{3.8}$$

If there is a q and a t_1 in $|t| > r$ for which the corresponding brace is not zero, then set $\{\Theta_q^{(2)}(u)\}$ is "linearly dependent" in $|u| < q_1$, contrary to hypothesis. Hence each brace is identically zero, thus making each $R_q^{(2)}$ a linear combination of the set $\{R_j^{(1)}\}$.

THEOREM 3.1. *Let* $\Phi(u, 1/t)$ *have the canonical form* (3.1) *relative to* $[q, q_1, q'; r]$, *with* $r < R$. *Then the homogeneous system* (1.2) $[c_n \equiv 0]$ *has the following* k *linearly independent solutions of type not exceeding* r:

$$\{x_n^{(j)}\} = \{r_{j,n}\} \qquad (j = 1, \cdots, k; n = 0, 1, \cdots), \tag{3.9}$$

where the r_{jn} *are defined by*

$$R_j(1/t) = \sum_{n=0}^{\infty} r_{jn}/t^n \qquad (j = 1, \cdots, k). \tag{3.10}$$

Proof. Let C be a circle, center $t = 0$ and radius between r and R. From (3.1),

$$\frac{1}{2\pi i} \int_C \Phi(u, 1/t) t^{n-1} A_n(t) \, dt = \sum_{j=1}^{k} \sigma_{jn} \Theta_j(u) + \sum_{j=0}^{\infty} \lambda_{jn} u^j, \tag{3.11}$$

valid for $|u| < q$. Moreover the last series converges in $|u| < q'$. Here

$$\sigma_{jn} = \frac{1}{2\pi i} \int_C t^{n-1} A_n(t) R_j(1/t) dt; \qquad \lambda_{jn} = \frac{1}{2\pi i} \int_C t^{n-1} A_n(t) H^*_j(1/t) dt. \tag{3.12}$$

Now for $|u| < q$, Φ also has the expansion (1.4), since $\mathfrak{a}$ satisfies $P_2[R, q, r]$; so for $|u| < q$ the left side of (3.11) has the value u^n. That is,

$$\sum_{j=1}^{k} \sigma_{jn} \Theta_j(u) = u^n - \sum_{j=0}^{\infty} \lambda_{jn} u^j \qquad (|u| < q). \tag{3.13}$$

But the right side is analytic in $|u| < q'$, and the set $\{\Theta_j\}$ is "linearly independent" in $|u| < q_1$. It follows that

$$(3.14) \qquad \sigma_{jn} = 0 \qquad (j = 1, \cdots, k; n = 0, 1, \cdots);$$

and from (3.12) and (2.6) we conclude that each sequence $\{r_{jn}\}$ $(j = 1, \cdots, k)$ is a solution of system (1.2) $[c_n \equiv 0]$.

Remarks. (i) Even if the functions $\{R_j(1/t)\}$ are not linearly independent the sequences $\{r_{jn}\}$ are solutions of system (1.2) $[c_n \equiv 0]$.

(ii) From (3.12), (3.13), (3.14) there follows the orthogonality relation

$$(3.15) \qquad \frac{1}{2\pi i}\int_C H^*_j(1/t)t^n A_n(t)\,dt/t = \delta_{jn},$$

where C is any rectifiable simple closed curve surrounding $t = 0$ and lying in the ring $r < |t| < R$.

It would be of interest to know if Theorem 3.1 supplies all the solutions of the homogeneous system, of type not exceeding r, without further hypotheses. Such knowledge is lacking, but with an additional condition the desired conclusion can be asserted.

THEOREM 3.2. *Let the hypotheses of Theorem* 3.1 *hold, as well as condition* $P_3[R, q_2, r_2]$ *with* $r < r_2$, $q_2 < q'$. *Then the only solutions of* (1.2) $[c_n \equiv 0]$, *of type less than* r_2, *are linear combinations of the* k *solutions of Theorem* 3.1.

Proof. Let $X = \{x_n\}$ be a solution of (1.2) $[c_n \equiv 0]$, of type $\sigma < r_2$. Let C be the circle $|t| = \rho$, with $\sigma < \rho < r_2$. From (2.6),

$$0 = \frac{1}{2\pi i}\int_C X(1/t)t^{n-1}A_n(t)\,dt.$$

Multiply through by u^{-n} and sum:

$$(3.16) \qquad 0 = \frac{1}{2\pi i}\int_C X(1/t)\Psi(1/u, t)\,dt/t \qquad (|u| > q_2).$$

Since the $\Phi^*(u, 1/s)$-series (3.2) is "uniformly" convergent for $|u| < q'$, $|s| > r$, we may multiply (3.16) through by $u^{-1}\Phi^*(u, 1/s)$ and integrate over the circle Γ: $|u| = q_3 (q_2 < q_3 < q')$:

$$(3.17) \qquad 0 = \frac{1}{2\pi i}\int_C X(1/t)\{\frac{1}{2\pi i}\int_\Gamma \Psi(1/u, t)\Phi^*(u, 1/s)\,du/u\}dt/t,$$

valid for $|s| > r$. Now for $|s| > r$, $|t| < r_2$ and u on Γ, the two series

(1. 5) and (3. 2), the latter with t replaced by s, converge "uniformly," so the brace, to be denoted by $U(t, 1/s)$, has the value

$$U(t, 1/s) = \sum_{n=0}^{\infty} t^n A_n(t) H^*{}_n(1/s), \tag{3.18}$$

"uniformly" convergent for $|t| < r_2$, $|s| > r$. Thus, (3. 17) becomes

$$0 = \frac{1}{2\pi i} \int_C X(1/t) U(t, 1/s)\, dt/t. \tag{3.19}$$

Now U can be expanded in a power series

$$U(t, 1/s) = \sum_{n=0}^{\infty} I_n(1/s) t^n, \tag{3.20}$$

"uniformly" convergent in $|t| < r_2$, $|s| > r$; so from (3. 19) and (3. 20),

$$0 = \sum_{n=0}^{\infty} x_n I_n(1/s), \tag{3.21}$$

"uniformly" convergent for $|s| > r$. If we substitute (3. 18 into

$$I_n(1/s) = \frac{1}{2\pi i} \int_C U(t, 1/s)\, dt/t^{n+1} \tag{3.22}$$

we find that for $|s| > r$,

$$I_n(1/s) = a_{0n} H^*{}_0(1/s) + a_{1,n-1} H^*{}_1(1/s) + \cdots + a_{n0} H^*{}_n(1/s). \tag{3.23}$$

From (3. 1), (3. 2) and (1. 4) we see that

$$H_n(1/t) = \sum_{j=1}^{k} \theta_{jn} R_j(1/t) + H^*{}_n(1/t), \tag{3.24}$$

valid for $|t| > r$, where

$$\Theta_j(u) = \sum_{n=0}^{\infty} \theta_{jn} u^n \qquad (j = 1, \cdots, k);$$

so

$$I_n(1/s) = \sum_{j=0}^{n} a_{j,n-j} H_j(1/s) - \sum_{p=0}^{n} a_{p,n-p} \{ \sum_{j=1}^{k} \theta_{jp} R_j(1/s) \}, \tag{3.25}$$

for $|s| > r$. Using (2. 2),

$$I_n(1/s) = 1/s^n - \sum_{j=1}^{k} \beta_{jn} R_j(1/s) \qquad (|s| > r) \tag{3.26}$$

where

$$\beta_{jn} = \sum_{p=0}^{n} a_{p,n-p} \theta_{jp}. \tag{3.27}$$

Returning to (3.21), we now have

$$(3.28)\qquad 0 = X(1/s) - \sum_{n=0}^{\infty} x_n\{\sum_{j=1}^{k} \beta_{jn}R_j(1/s)\}\qquad (|s| > \max\{\sigma, r\}).$$

We wish to interchange the summation signs, but this requires justification. Since $\{R_j\}$ is a linearly independent set, given an arbitrary number $r' > r$, there exist k values $s = s_1, \cdots, s_k$ in $|s| > r'$ such that the determinant $|R_j(1/s_p)|$ $(j, p = 1, \cdots, k)$ has a value $\Delta \neq 0$. Let C_{jp} be the cofactor of $R_j(1/s_p)$ in this determinant. Then from (3.28),

$$\sum_{p=1}^{k} C_{qp}X(1/s_p) = \sum_{n=0}^{\infty} x_n\{\sum_{j=1}^{k} \beta_{jn}(\sum_{p=1}^{k} C_{q,p}R_j(1/s_p))\}.$$

Now the last sum on the right has the value $\delta_{qj}\Delta$, where δ_{qj} is the Kronecker symbol; so

$$(3.29)\qquad \sum_{n=0}^{\infty} x_n\beta_{qn} = 1/\Delta \sum_{p=1}^{k} C_{qp}X(1/s_p).$$

Thus, for each $q = 1, \cdots, k$ the series on the left of (3.29) converges, and may be multiplied by $R_q(1/s)$ and summed from $q = 1$ to k, the resulting series being valid for $|s| > \sigma$. The double sum in (3.28) may therefore be rearranged to give the relation

$$(3.30)\qquad X(1/s) = \sum_{j=1}^{k} \lambda_j R_j(1/s)$$

where $\lambda_j = \sum_{n=0}^{\infty} x_n\beta_{jn}$. This establishes the theorem.

There is an extension of Theorem 3.1 to the case of infinitely many solutions.

Definition. The infinite sequence of functions $\{\Theta_j(u)\}$, $j = 0, 1, \cdots$ is "linearly dependent" in $|u| < q$ if constants $\{a_j\}$ exist, not all zero, such that series $\sum_{0}^{\infty} a_j\Theta_j(u)$ converges "uniformly" in a circle $|u| < \rho$ with $\rho > q$.

Definition. $\Phi(u, 1/t)$ is said to have the canonical representation

$$(3.31)\qquad \Phi(u, 1/t) = \sum_{j=0}^{\infty} R_j(1/t)\Theta_j(u) + \Phi^*(u, 1/t)$$

relative to $[q, q_1, q'; r]$ if: i) all the conditions of the earlier definitions[5] (with respect to (3.1)) hold except the requirement that $\{R_j(1/t)\}$ be linearly independent; and ii) the series $\sum_{0}^{\infty} R_j(1/t)\Theta_j(u)$ is "uniformly" convergent in $|u| < q$, $|t| > r$.

---

[5] Of course, the new definition of "linear independence" is to be used.

THEOREM 3.3. *If* $\Phi(u, 1/t)$ *has the canonical representation* (3.31) *relative to* $[q, q_1, q'; r]$, *then every sequence* $\{r_{jn}\}$, $j = 0, 1, \cdots$ *is a solution of system* (1.2) $[c_n \equiv 0]$, *of type not exceeding* r.

Proof. As with Theorem 3.1.

There is also a parallel with Theorem 3.2.

THEOREM 3.4. *Let the hypotheses of Theorem* 3.3 *hold, and also condition* $P_3[R, q_2, r_2]$ *with* $r < r_2$, $q_2 < q'$. *There exists an infinite sequence of solutions* $\{r^*_{jn}\}$, *of type not exceeding* r, *such that on setting*

$$(3.32) \qquad R^*_j(1/t) = \sum_{n=0}^{\infty} r^*_{jn}/t^n \qquad (j = 0, 1, \cdots),$$

then every solution $\{x_n\}$ *of type less than* r_2 *is given by the representation*

$$(3.33) \qquad X(1/t) = \sum_{j=0}^{\infty} x_j R^*_j(1/t),$$

"uniformly" convergent in $|t| > \max\{r, \sigma\}$ *where* $\sigma = ((x_n))$.

Proof. As with Theorem 3.2, through (3.28) with the sum $j = 1, \cdots, k$ replaced by $j = 0$ to ∞, to yield

$$(3.34) \qquad 0 = X(1/s) - \sum_{n=0}^{\infty} x_n \{ \sum_{j=0}^{\infty} \beta_{jn} R_j(1/s) \},$$

"uniformly" convergent in $|s| > \max\{r, \sigma\}$. The β_{jn} are still given by (3.27). Now define $R^*_n(1/t)$ by

$$(3.35) \qquad R^*_n(1/t) = \sum_{j=0}^{\infty} \beta_{jn} R_j(1/t), \qquad n = 0, 1, \cdots.$$

This series is "uniformly" convergent in $|t| > r$, so $R^*_j(1/t)$ is analytic in $|t| > r$ and $((r^*_{jn})) \leqq r$ $(j = 0, 1, \cdots)$. Also, since each $\{r_{jn}\}$ is a solution of (1.2) $[c_n \equiv 0]$, we can apply relation (2.6) to (3.35) to show that each $\{r^*_{jn}\}$ is also a solution. The theorem now follows from (3.34).

Remarks. (i) It is of interest to observe that the coefficients $\{x_n\}$ in the expansion (3.33) are also the coefficients in the power series (in $1/t$) for $X(1/t)$.

(ii) Theorems 3.1 and 3.2 also have parallels in the continuous case. Suppose, for example, that

$$\Phi(u, 1/t) = \int_a^b R(1/t, \omega) \Theta(u, w) \, d\alpha(\omega) + \Phi^*(u, 1/t),$$

where Φ^* is as before. Assume that $\alpha(\omega)$ is of bounded variation on $I: [a, b]$, and that R, Θ are continuous for $|t| > r$, ω on I, and $|u| < q$, ω on I, respectively, and for each ω on I, the functions R, Θ are analytic in $|t| > r$, $|u| < q$ respectively. Also suppose that Θ is "linearly independent" in $|u| < q_1$, in the sense that if $a(\omega)$ is continuous on I and $\int_a^b a(\omega)\Theta(u, \omega)d\alpha(\omega)$ is analytic in $|u| < \rho$ with $\rho > q_1$, then $a(\omega) \equiv 0$. It will then follow that for each ω on I the sequence $r_n(\omega)$ defined by

$$R(1/t, \omega) = \sum_{n=0}^{\infty} r_n(\omega)/t^n$$

is a solution of (1.2) $[c_n \equiv 0]$, and is of type not exceeding r. And, as in Theorem 3.2 (with the appropriate condition P_3), if $\{x_n\}$ is a solution of type σ less than r, then $X(1/t)$ has the representation

$$X(1/t) = \sum_{n=0}^{\infty} x_n \{ \int_a^b \beta_n(\omega) R(1/t, \omega) d\alpha(\omega),$$

"uniformly" convergent in $|t| > \max\{r, \sigma\}$, where

$$\beta_n(\omega) = \sum_{p=0}^{n} a_{p,n-p}\theta_p(\omega), \qquad \Theta(u, \omega) = \sum_{p=0}^{\infty} \theta_p(\omega) u^p.$$

For the non-homogeneous system we have the following extension of Theorem 2.3.

THEOREM 3.5. *Let the hypotheses of Theorem 3.1 hold. For every sequence $\{c_n\}$ with $((c_n)) < \min\{q', R\}$, system (1.2) has the solution $\{x_n\}$ given by*

(3.36) $$X(1/t) = \frac{1}{2\pi i}\int_\Gamma \Phi^*(u, 1/t) C(1/u) du/u,$$

Γ being a rectifiable simple closed curve surrounding $u = 0$ and lying in the ring $((c_n)) < |u| < q'$. Moreover, $((x_n)) \leqq r$.

Proof. Using (2.6),

(3.37) $$A_n[X] = \frac{1}{2\pi i}\int_\Gamma C(1/u) V_n(u) du/u,$$

where

(3.38) $$V_n(w) = \frac{1}{2\pi i}\int_C \Phi^*(u, 1/t) t^{n-1} A_n(t) dt$$

and C is the circle $|u| = \rho$, $r < \rho < R$. The function $V_n(u)$ is analytic in $|u| < q'$, and from (3.15) its value is u^n; so $A_n[X] = c_n$, $n = 0, 1, \cdots$. That $((x_n)) \leqq r$ is clear from (3.36).

Remarks. (i) A similar result holds if we assume the hypotheses of Theorem 3. 3 or of the second remark following Theorem 3. 4.

(ii) The solution (3. 36) also has the representation

$$X(1/t) = \sum_{n=0}^{\infty} c_n H^*_n(1/t), \tag{3.39}$$

the series being "uniformly" convergent in $|t| > r$; and it is readily shown that conversely, if $\{c_n\}$ is such that (3. 39) converges "uniformly" in $|t| > \delta$ for some $\delta < R$, then $\{x_n\}$ given by $X(1/t) = \Sigma x_n/t^n$ is a solution of (1. 2) of type not exceeding δ. (Remark (i) also applies here.)

If $\Phi(u, 1/t)$ does not possess a canonical form (3. 1) or the later extensions thereof, it is yet possible in certain cases to extend Theorem 2. 3, as the following concluding result indicates.

THEOREM 3. 6. *Let $P_2[R, q, r]$ hold with $r < R$. Let there exist a rectifiable simple closed curve γ surrounding $u = 0$ and lying in $|u| > q$, with the following two properties:* i) *There is an open connected set $\mathcal{S}$ cotnaining γ and the point $w = 0$, in which the functions $L_n(u)$ of* (2. 12) *are single-valued and regular;* ii) *$r_1 < R$ exists such that series* (2. 12) *converges "uniformly" for u on $\mathcal{S}$ and $|t| > r_1$. Let q_1 be the minimum distance from γ to $u = 0$. Then for every $\{c_n\}$ for which $((c_n)) < q_1$, system* (1. 2) *has the solution $\{x_n\}$ given by* (2. 14). *Moreover, $((x_n)) \leqq r_1$.*

Proof. Choose r_2 so that $r < r_2 < R$. There is an M such that $|\Phi(u, 1/t)| \leqq M$ for u on γ and $|t| = r_2$, so for u on γ, $|u^n L_n(u)| \leqq M r_2{}^n$. Let $\{x_n\}$ be defined by (2. 14). Then $((x_n)) \leqq r_2$, and since r_2 can be chosen as close to r_1 as we please, therefore $((x_n)) \leqq r_1$. Also, for each $n = 0, 1, \cdots$,

$$A_n[X] = \sum_{j=0}^{\infty} a_{nj} \{ \frac{1}{2\pi i} \int_\gamma C(1/u) u^{n+j-1} L_{n+j}(u)\, du \}.$$

The integral and sum are readily seen to be interchangeable, so

$$A_n[X] = \frac{1}{2\pi i} \int_\gamma C(1/u) \{ \sum_{j=0}^{\infty} a_{nj} \cdot \frac{1}{2\pi i} \int_{|t|=r_2} \Phi(u, 1/t) t^{n+j-1} dt \} du/u$$

$$= \frac{1}{2\pi i} \int_\gamma C(1/u) \{ \frac{1}{2\pi i} \int_{|t|=r_2} t^{n-1} \Phi(u, 1/t) A_n(t)\, dt \} du/u.$$

For $|u| < q$, and taking r_2 in $r < r_2 < R$, we see from (1. 4) that the brace has the value u^n. But by hypothesis the brace is single-valued and regular in $\mathcal{S}$, so its value remains u^n in $\mathcal{S}$, and in particular on γ. Thus, $A_n[X] = c_n$; that is, $\{x_n\}$ is a solution of system (1. 2).

THE PENNSYLVANIA STATE COLLEGE.

BIBLIOGRAPHY.

[1] R. P. Boas, Jr., "Basic sets of polynomials. II," *Duke Journal of Mathematics*, vol. 16 (1949), pp. 145-149.

[2] J. Horn, "Volterrasche Integralgleichungen und Summenleichungen," *Journal für reine und angewandte Mathematische*, vol. 140 (1911), pp. 120-174.

[3] ———, "Analytische Lösungen von Summengleichungen," *Archiv der Mathematik und Phys.*, ser. 3, vol. 26 (1918), pp. 132-145.

[4] O. Perron, "Zur Theorie der Summengleichungen," *Mathematische Zeitschrift*, vol. 8 (1920), pp. 159-170.

[5] ———, "Über Summengleichungen und Poincarésche Differengleichungen," *Mathematische Annalen*, vol. 84 (1921), pp. 1-15.

[6] I. M. Sheffer, "An extension of a Perron system of linear equations in infinitely many unknowns," *American Journal of Mathematics*, vol. 67 (1945), pp. 123-140.

[7] ———, "On k-periodic systems of linear equations," *Transactions of the American Mathematical Society*, vol. 63 (1948), pp. 244-313.

[8] ———, "On the theory of sum-equations," *Bulletin of the American Mathematical Society*, vol. 55 (1949), pp. 777-788.

[9] H. von Koch, "On regular and irregular solutions of some infinite systems of linear equations," *Proceedings of the Fifth International Congress of Mathematicians*, Cambridge (1912), vol. 1, pp. 352-365.

[10] J. L. Walsh, "On the solution of linear equations in infinitely many variables by successive approximations," *American Journal of Mathematics*, vol. 42 (1920), pp. 91-96.

THE GAPS IN THE ESSENTIAL SPECTRA OF WAVE EQUATIONS.*

By Philip Hartman and Calvin R. Putnam.

1. Let $q = q(t)$, where $0 \leqq t < \infty$, be a continuous, real-valued function for which the differential equation

$$x'' + (\lambda + q)x = 0 \tag{1}$$

is of the Grenzpunkt type; [9], p. 238. A fixed homogeneous boundary condition at $t = 0$,

$$x(0) \cos \alpha + x'(0) \sin \alpha = 0, \tag{2}$$

determines a self-adjoint boundary value problem with a (real) spectrum $S = S(\alpha)$. The set S' of cluster points of $S(\alpha)$ is independent of α ([9], p. 251) and will be called the essential spectrum of (1).

Since S' is a closed set, its complement is open, possibly empty. Let $\Sigma(\lambda_k, \lambda^k)$ denote the canonical decomposition of the complement of S' into disjoint open intervals. Two problems will be treated in this paper. The first concerns the size of the interval (λ_k, λ^k) as $\lambda_k \rightarrow \infty$. The second concerns asymptotic estimates, as $t \rightarrow \infty$, for eigenfunctions $x = x(t, \lambda)$ belonging to eigenvalues λ of $S(\alpha)$ satisfying $\lambda_k < \lambda < \lambda^k$ for a large value of λ_k. These problems were suggested to us by Professor Wintner.

The nature of the problems is most easily understood in the case that $q(t)$ is periodic. For this case, let $L(\lambda) \geqq 0$ denote the real part of one of the characteristic exponents belonging to λ. Then $L(\lambda) = 0$ except on a half-line $(-\infty, \lambda^0)$ and a sequence of intervals (λ_1, λ^1), $(\lambda_2, \lambda^2), \cdots$, composed of the λ-values for which (1) is of unstable type. The questions treated concern upper estimates for the length of the interval (λ_k, λ^k), as $\lambda_k \rightarrow \infty$, and upper estimates for $L(\lambda)$, as $\lambda \rightarrow \infty$. The results below concerning $L(\lambda)$, which imply that $L(\lambda) \rightarrow 0$ as $\lambda \rightarrow \infty$, show that, as λ increases, any tendency for (1) to be unstable is dampened; in the sense, that if λ is large and of unstable type, it requires long t-intervals in order that the effect of the factor e^{Lt} be appreciable.

In the case that $q(t)$ is even and periodic, the behavior of $\lambda^k - \lambda_k$ can be read off from known results on the asymptotic behavior of eigenvalues of Sturm-Liouville boundary value problems; cf. Section 7, below. This pro-

---

* Received March 23, 1950.

cedure can be adapted also for the case that $q(t)$ is merely periodic, not necessarily even, if it is noted that the k-th interval $[\lambda_k, \lambda^k]$ is composed of those λ-values corresponding to which (1) has a solution $x = x(t)$ which vanishes exactly k times for $0 \leqq t < p$, satisfies (2) and the same boundary condition $x(p) \cos \alpha + x'(p) \sin \alpha = 0$ at $t = p$ for some $0 \leqq \alpha < \pi$, when p is the period of $q(t)$.

Of course, the results of this paper are not restricted to the cases where $q(t)$ is periodic.

2. The gaps (λ_k, λ^k) in the essential spectrum. A sufficient condition for (1) to be of the Grenzpunkt type is that $\beta < \infty$, where $\beta = \limsup q(t)$, as $t \to \infty$ ([9], p. 238). This sufficient condition does not involve $\alpha = \liminf q(t) \geqq -\infty$, as $t \to \infty$. If $q(t)$ tends to a finite limit as $t \to \infty$ (that is, $-\infty < \alpha = \beta < \infty$), then the essential spectrum S' of (1) contains the half-line $\beta \leqq \lambda < \infty$; [3], p. 71. More generally, if $q(t)$ is bounded (that is, $-\infty < \alpha \leqq \beta < \infty$), then, for every $\lambda_0 \geqq -\beta$, the interval $\lambda_0 \leqq \lambda \leqq \lambda_0 + \beta - \alpha$ contains at least one point of S'; [6], p. 613. The object of the first part of this paper will be the refinement of this last statement in the following direction:

(I) *Let* $q(t)$, *where* $0 \leqq t < \infty$, *be a real-valued, continuous, bounded function; for* $\delta > 0$, *let*

$$\omega(t, \delta) = \text{fin sup} \mid q(s) - q(t) \mid \text{ for } \mid s - t \mid \leqq \delta; \tag{3}$$

and

$$\omega(\delta) = \limsup T^{-1} \int_0^T \omega(t, \delta)\, dt \text{ as } T \to \infty. \tag{4}$$

Then there exists a constant C *(independent of* λ_0*) such that, for every sufficiently large* λ_0, *the interval* $\lambda_0 \leqq \lambda \leqq \lambda_0 + C[\omega(\pi/\lambda_0^{\frac{1}{2}}) + 1/\lambda_0]$ *contains at least one point of the essential spectrum* S' *of* (1).

This theorem has the following interpretation and consequences: Since q is bounded, the complement of S' consists of a half-line $(-\infty, \lambda^0)$ and a (possibly empty) sequence of disjoint open intervals (λ_1, λ^1), $(\lambda_2, \lambda^2), \cdots$. The theorem (I) states that the length $\lambda^k - \lambda_k$ of the k-th gap is majorized by $C[\omega(\pi/\lambda_k^{\frac{1}{2}}) + 1/\lambda_k]$ for large λ_k. In particular, if $q(t)$ is uniformly continuous for $0 \leqq t < \infty$, so that $\omega(\delta) \to 0$ as $\delta \to +0$, then the length of the gap tends to 0 as its endpoints tend to ∞.

Similarly, if $q(t)$ satisfies a uniform Lipschitz condition, $|q(s)-q(t)| \leqq \text{const.}\,|s-t|$ for $0 \leqq s < t < \infty$, so that $\omega(\delta) \leqq \text{const.}\,\delta$, then

$$\lambda^k - \lambda_k = O(1/\lambda_k^{\frac{1}{2}}). \tag{5}$$

This same majorant for $\lambda^k - \lambda_k$ is obtained if $q(t)$ satisfies the more general conditions that $q(t)$ is of bounded variation on every finite interval and that its total variation $V(T)$ on the interval $0 \leqq t \leqq T$ satisfies

$$V(T) = \int_0^T |dq(t)| = O(T) \text{ as } T \to \infty. \tag{6}$$

In order to see this, note that $\omega(t,\delta) \leqq V(t+\delta) - V(t-\delta)$ for $t \geqq \delta$, by (3). Hence, if $T \geqq \delta$,

$$\int_\delta^T \omega(t,\delta)\,dt \leqq \int_{2\delta}^{T+\delta} V(t)\,dt - \int_0^{T-\delta} V(t)\,dt \leqq \int_{T-\delta}^{T+\delta} V(t)\,dt.$$

Then, since $V(t) \leqq \text{Const.}\,t$, the right side of the last inequality is majorized by Const. δT, as $T \to \infty$, for $\delta > 0$ fixed. Consequently, by (4), it is seen that $\omega(\delta) \leqq \text{const.}\,\delta$ and so the majorant $C/\lambda_k^{\frac{1}{2}}$ of $\lambda^k - \lambda_k$ for large λ_k follows from (I).

In the few explicit cases where q is bounded and S' is known, either there is only a finite number of gaps, or $\lambda^k - \lambda_k \to 0$ and $\lambda_k \to \infty$, as $k \to \infty$. Whether or not this is true in general will remain an open question. On the other hand, if $\psi(\lambda)$ is an arbitrary decreasing function satisfying $\psi(\lambda) \to 0$ as $\lambda \to \infty$, then there exist bounded functions q, corresponding to which the complement of the essential spectrum S' of (1) consists of a half-line $(-\infty, \lambda^0)$, and a sequence of open intervals (λ_1, λ^1), $(\lambda_2, \lambda^2), \cdots$ such that $\lambda_k \to \infty$, as $k \to \infty$, but $\lambda^k - \lambda_k = O(\psi(\lambda_k))$ fails to hold; see Section [illegible] below. Also, examples can be given to show that if $q(t)$ satisfies (6), then (5) is the best possible estimate; in the sense that, in general, O cannot be replaced by o in (5).

If additional local smoothness conditions are imposed on q, the assertions of Theorem (I) can be refined.

(I′) *Let* $q(t)$ *satisfy the conditions of* (I) *and; in addition, possess a bounded derivative* $q'(t)$. *If* $\omega^*(t,\delta)$ *and* $\omega^*(\delta)$ *have the same significance for* q' *as* (3) *and* (4) *do for* q, *then there exists a constant* C (*independent of* λ_0) *such that, for every sufficiently large* λ_0, *the interval* $\lambda_0 \leqq \lambda \leqq \lambda_0 + C\lambda_0^{-\frac{1}{2}}[\omega^*(\pi/\lambda_0^{\frac{1}{2}}) + 1/\lambda_0]$ *contains at least one point of the essential spectrum* S' *of* (1).

(I$''$) *Furthermore, if* $q(t)$ *possesses a uniformly bounded first and second derivative and if* $\omega^{**}(t,\delta)$, $\omega^{**}(\delta)$ *belong to* q'' *as* (3), (4) *do to* q, *then the assertion of* (I$'$) *applies to intervals* $\lambda_0 \leqq \lambda \leqq \lambda_0 + C\lambda_0^{-1}[\omega^{**}(\pi/\lambda_0^{\frac{1}{2}}) + 1/\lambda_0^{\frac{1}{2}}]$, *for some constant* C *and all sufficiently large* λ_0.

Thus, it follows from (I$'$) that if $q'(t)$ satisfies a uniform Lipschitz condition, or more generally, if

$$\int_0^T |\,dq'(t)\,| = O(T),$$

as $T \rightarrow \infty$, then the length of the k-th gap (λ_k, λ^k) is majorized by $O(1/\lambda_k)$ as $\lambda_k \rightarrow \infty$. In addition, the existence of a bounded second derivative satisfying

$$\int_0^T |\,dq''(t)\,| = O(T),$$

as $T \rightarrow \infty$, implies that $\lambda^k - \lambda_k = O(1/\lambda_k^{3/2})$ as $\lambda_k \rightarrow \infty$.

3. The characteristic exponents. If $q(t)$ denotes a continuous, bounded, real-valued function, then it is known, [11], p. 391, that there exists a positive number $k = k(\lambda)$, depending on λ alone, such that

$$x^2 + x'^2 = O(e^{kt}), \tag{7}$$

as $t \rightarrow \infty$, holds for any solution $x = x(t,\lambda)$ of the differential equation (1). Let the number $K = K(\lambda)$ be defined as the greatest lower bound of the numbers k for which (7) holds (for *every* solution x of (1)). As shown in [7], p. 137, in partial answer to a question raised by Wintner [14], p. 604, if $|\,q(t)\,|$ possesses the bound M and if λ does not belong to the set S' and satisfies $|\,\lambda\,| > M$, then there exist a constant $l = l(\lambda) > 0$ and two solutions $x = x_1(t,\lambda) \not\equiv 0$ and $x = x_2(t,\lambda)$ of (1) satisfying

$$x_1^2 + x_1'^2 \leqq e^{-lt}, \qquad x_2^2 + x_2'^2 \geqq e^{lt}. \tag{8}$$

If $L = L(\lambda)$ is defined to be the least upper bound of the numbers l for which (8) holds, then it is clear from the definition of $K(\lambda)$ that $L(\lambda) \leqq K(\lambda)$, and hence any majorant of $K(\lambda)$ will also be a majorant of $L(\lambda)$. Since (1) is of the Grenzpunkt type, the solution $x_1(t,\lambda)$ is uniquely determined to within a constant factor. Thus, it is clear that $L(\lambda)$ depends on λ alone and not on the choice of the particular pair of solutions x_1 and x_2 of (1). The number $L(\lambda)$ can be called a "characteristic exponent" associated with

λ, in analogy with the terminology used in the case that $q(t)$ is periodic. It will be the purpose of the second part of the paper to obtain, for large λ, estimates of $K(\lambda)$, and hence of $L(\lambda)$, when $q(t)$ is subject to certain smoothness conditions.

In Sections **8** and **9**, the following theorems will be proved:

(II) *If* $q(t)$ *satisfies the conditions of* (I), *then*

$$K(\lambda) = O(\lambda^{-\frac{1}{2}}[\omega(\pi/\lambda^{\frac{1}{2}}) + 1/\lambda]) \text{ as } \lambda \to \infty. \tag{9}$$

As a consequence of (II) it is seen that, since $q(t)$ is bounded, the estimate $K(\lambda) = O(\lambda^{-\frac{1}{2}})$ is obtained. If, in addition, $q(t)$ is uniformly continuous for $0 \leqq t < \infty$, the last estimate can be sharpened to $K(\lambda) = o(\lambda^{-\frac{1}{2}})$. Also, if $q(t)$ satisfies (6), then $K(\lambda) = O(\lambda^{-1})$. Further smoothness conditions imply better estimates for $K(\lambda)$; in this direction the following will be proved:

(II′) *If* $q(t)$ *satisfies the conditions of* (I′), *then*

$$K(\lambda) = O(\lambda^{-1}[\omega^{*}(\pi/\lambda^{\frac{1}{2}}) + 1/\lambda]) \text{ as } \lambda \to \infty. \tag{10}$$

(II″) *If* $q(t)$ *satisfies the conditions of* (I″), *then*

$$K(\lambda) = O(\lambda^{-3/2}[\omega^{**}(\pi/\lambda^{\frac{1}{2}}) + 1/\lambda^{\frac{1}{2}}]) \text{ as } \lambda \to \infty. \tag{11}$$

4. Proof of (I). Let $x = x(t, \lambda) \not\equiv 0$ be a solution of (1) and let $N(T, \lambda)$ denote the number of zeros of x for $0 < t < T$. According to [4], a λ-interval $(\lambda, \lambda + \Delta)$ contains a point of S' if and only if

$$N(T, \lambda + \Delta) - N(T, \lambda) \to \infty \text{ as } T \to \infty. \tag{12}$$

Since it is assumed that q is bounded, there exists a constant M such that

$$|q| \leqq M, \qquad (0 \leqq t < \infty). \tag{13}$$

Let λ be chosen so large that

$$\lambda - M > \lambda/4 > 0. \tag{14}$$

Then the zeros of x and x' separate each other. Hence, if $\theta = \theta(t, \lambda)$ is a continuous determination of

$$\theta(t, \lambda) = -\operatorname{arc\,tan}(x'/\lambda^{\frac{1}{2}}x), \tag{15}$$

it follows that $\pi N(T, \lambda) = \theta(T, \lambda) + O(1)$, as $T \to \infty$. Consequently, (12) holds if and only if

$$\theta(T, \lambda + \Delta) - \theta(T, \lambda) \to \infty \text{ as } T \to \infty. \tag{16}$$

It follows from (1) and from (15) that θ is differentiable and that $\theta' = \lambda^{\frac{1}{2}} + \lambda^{-\frac{1}{2}} q\{\lambda x^2/(\lambda x^2 + x'^2)\}$; so that

$$(17) \qquad \theta(T, \lambda) = \lambda^{\frac{1}{2}} T + \lambda^{-\frac{1}{2}} \int_0^T q\{\lambda x^2/(\lambda x^2 + x'^2)\} dt + \theta(0, \lambda).$$

Let $(0 \leqq) t_1 < t_2 < \cdots$ be the zeros of x. Then

$$(18) \qquad t_{k+1} - t_k \leqq 2\pi/\lambda^{\frac{1}{2}},$$

by virtue of Sturm's comparison theorem and (14). The factor $\{\cdots\}$ of q in the integral in (17) will be estimated on the interval (t_k, t_{k+1}). Because of the homogeneity of this factor, x can be replaced by any solution y of (1) which vanishes for $t = t_k$.

Let ν_k, where $k = 1, 2, \cdots$, be defined by

$$(19) \qquad t_{k+1} - t_k = \pi/\nu_k^{\frac{1}{2}}.$$

Clearly, $\nu_k = \lambda + q(\xi_k)$, where ν_k is some t-value satisfying $t_k \leqq \xi_k \leqq t_{k+1}$. Hence,

$$(20) \qquad |\lambda - \nu_k| \leqq M.$$

A solution y of (1) which vanishes at $t = t_k$ satisfies the integral equation

$$(21) \qquad y(t) = \sin \nu_k^{\frac{1}{2}}(t - t_k) - \nu_k^{-\frac{1}{2}} \int_{t_k}^{t} Q_k(s) y(s) \sin \nu_k^{\frac{1}{2}}(t - s)\, ds,$$

where

$$(22) \qquad Q_k(s) = q(s) - q(\xi_k).$$

Let $m = m_k = \max |y(t)|$ for $t_k \leqq t \leqq t_{k+1}$; then, by (13) and (19), it [illegible]llows that $m \leqq 1 + \nu_k^{-\frac{1}{2}}(\pi \nu_k^{-\frac{1}{2}}) 2Mm$. Consequently, $m \leqq 1/(1 - 2\pi M \nu_k^{-1}) \leqq 2$ [illegible]f ν_k (that is, λ) is sufficiently large. Thus (21) shows that

$$(23) \qquad y = \sin \nu_k^{\frac{1}{2}}(t - t_k) + O(\lambda^{-1})$$

and

$$(24) \qquad y' = \nu_k^{\frac{1}{2}} \cos \nu_k^{\frac{1}{2}}(t - t_k) + O(\lambda^{-\frac{1}{2}})$$

uniformly in k, as $\lambda \to \infty$. Since (20) implies that $|\nu_k^{\frac{1}{2}} - \lambda^{\frac{1}{2}}| = O(\lambda^{-\frac{1}{2}})$, it follows from (24) that

$$(25) \qquad y' = \lambda^{\frac{1}{2}} \cos \nu_k^{\frac{1}{2}}(t - t_k) + O(\lambda^{-\frac{1}{2}}).$$

Hence, for $t_k \leqq t \leqq t_{k+1}$,

$$(26) \qquad \lambda y^2/(\lambda y^2 + y'^2) = \sin^2 \nu_k^{\frac{1}{2}}(t - t_k) + O(\lambda^{-1}),$$

as $\lambda \to \infty$. Consequently, (17) shows that, for every fixed sufficiently large λ,

$$(27) \quad \theta(T,\lambda) = \lambda^{\frac{1}{2}}T + \tfrac{1}{2}\lambda^{-\frac{1}{2}} \int_0^T q(s)\,ds - \tfrac{1}{2}\lambda^{-\frac{1}{2}} \sum_{k=1}^{n} \int_{t_k}^{t_{k+1}} q(s) \cos 2\nu_k^{\frac{1}{2}}(s - t_k)\,ds + O(\lambda^{-3/2}T),$$

as $T \to \infty$, where $n = n(T)$ is determined by $t_n \leqq T < t_{n+1}$. In (27) and similar relations below, the constant implicit in the O-term can be chosen independent of λ.

Since the definition (19) of ν_k implies

$$\int_{t_k}^{t_{k+1}} \cos 2\nu_k^{\frac{1}{2}}(s - t_k)\,ds = 0 \text{ for } k = 1, 2, \cdots,$$

it follows that in the integral over the interval (t_k, t_{k+1}) in (27), the function $q(s)$ can be replaced by $q(s) - c_k$, where $c_1, c_2, \cdots$ are arbitrary constants. If $c_k = q(\tfrac{1}{2}t_k + \tfrac{1}{2}t_{k+1})$, then

$$| q(s) - c_k | \leqq \omega(s, \pi/\lambda^{\frac{1}{2}}), \text{ for } t_k \leqq s \leqq t_{k+1},$$

by (3) and (18). Hence,

$$\theta(T,\lambda) = \lambda^{\frac{1}{2}}T + \tfrac{1}{2}\lambda^{-\frac{1}{2}} \int_0^T q(s)\,ds + O(\lambda^{-\frac{1}{2}}) \int_0^T \omega(s, \pi/\lambda^{\frac{1}{2}})\,ds + O(\lambda^{-3/2}T).$$

Then, since $(\lambda + \Delta)^{\frac{1}{2}} - \lambda^{\frac{1}{2}} > \Delta\lambda^{-\frac{1}{2}}/4$ for any fixed Δ and for sufficiently large λ, and since the first integral of the last formula line is $O(T)$ as $T \to \infty$, it is clear that

$$\theta(T, \lambda + \Delta) - \theta(T,\lambda) \geqq \Delta(\lambda^{-\frac{1}{2}}T/4 + O(\lambda^{-3/2}T) + O(\lambda^{-\frac{1}{2}}) \int_0^T \omega(s, \pi/\lambda^{\frac{1}{2}})\,ds + \text{[illegible]}(\lambda^{-3/$$

as $T \to \infty$; so that there exists a constant C with the property that if $\Delta \geqq C\omega(\pi/\lambda^{\frac{1}{2}}) + C/\lambda$, then (16) holds for sufficiently large λ. This proves (I).

Incidentally, since the factor 1/4 on the right of (14) can be replaced by $1 - \epsilon$, where $\epsilon > 0$ is arbitrary, it follows that $\omega(\pi/\lambda^{\frac{1}{2}})$ can be replaced by $\omega(\pi/2(1-\epsilon)^{\frac{1}{2}}\lambda^{\frac{1}{2}})$ in the statement of the theorem (I).

5. Proof of (I′). The proof depends on a refinement of (26). The existence of a bounded derivative q' and (22) show that $Q_k(s) = O(\lambda^{-\frac{1}{2}})$, as $\lambda \to \infty$, uniformly for $t_k \leqq s \leqq t_{k+1}$. Hence, the remainder terms in (23) and (24) are $O(\lambda^{-3/2})$ and $O(\lambda^{-1})$, respectively. Thus

$$y'^2 = \nu_k \cos^2 \nu_k^{\frac{1}{2}}(t - t_k) + O(\lambda^{-\frac{1}{2}}),$$

or, since $\nu_k = \lambda + q(\xi_k) = \lambda + q(t) + O(\lambda^{-\frac{1}{2}})$,

$$y'^2\lambda^{-1} = \cos^2 \nu_k^{\frac{1}{2}}(t - t_k) + q(t)\lambda^{-1}\cos^2 \nu_k^{\frac{1}{2}}(t - t_k) + O(\lambda^{-3/2}).$$

Hence,

$$y^2 + y'^2\lambda^{-1} = 1 + q(t)\lambda^{-1}\cos^2 \nu_k^{\frac{1}{2}}(t - t_k) + O(\lambda^{-3/2}).$$

Finally, the required refinement of (26) is

$$(28)\quad \lambda y^2/(\lambda y^2 + y'^2) = \sin^2 \nu_k^{\frac{1}{2}}(t - t_k) \\ - q(t)(4\lambda)^{-1}\sin^2 2\nu_k^{\frac{1}{2}}(t - t_k) + O(\lambda^{-3/2}),$$

as $\lambda \to \infty$, uniformly for $t_k \leqq t \leqq t_{k+1}$.

It follows from (17) that, for fixed large λ, the function $\theta(T, \lambda)$ is the sum of the three expressions

$$(29)\qquad \lambda^{\frac{1}{2}}T + \tfrac{1}{2}\lambda^{-\frac{1}{2}}\int_0^T q(t)\,dt - (1/8\lambda^{3/2})\int_0^T q^2(t)\,dt,$$

$$(30)\qquad -\tfrac{1}{2}\lambda^{-\frac{1}{2}}\sum_{k=1}^{n}\int_{t_k}^{t_{k+1}} q(t)\cos 2\nu_k^{\frac{1}{2}}(t - t_k)\,dt,$$

$$(31)\qquad + (1/8\lambda^{3/2})\sum_{k=1}^{n}\int_{t_k}^{t_{k+1}} q^2(t)\cos 4\nu_k^{\frac{1}{2}}(t - t_k)\,dt + O(\lambda^{-2}T),$$

[illegible] $T \to \infty$, where $n = n(T)$ satisfies $t_n \leqq T < t_{n+1}$. An integration by parts and (19) show that the k-th term of the sum in (30) is

$$(32)\qquad -\tfrac{1}{2}\nu_k^{-\frac{1}{2}}\int_{t_k}^{t_{k+1}} q'(t)\sin 2\nu_k^{\frac{1}{2}}(t - t_k)\,dt;$$

since $q'(t)$ can be replaced by $q'(t) - c$, where c is an arbitrary constant, it follows from the choice $c = q'(\tfrac{1}{2}t_k + \tfrac{1}{2}t_{k+1})$, that the expression (32) is

$$O(\lambda^{-\frac{1}{2}})\int_{t_k}^{t_{k+1}} \omega^*(t, \pi/\lambda^{\frac{1}{2}})\,dt.$$

Hence, (30) is

$$O(\lambda^{-1})\int_0^T \omega^*(t,\pi/\lambda^{\frac12})\,dt+O(\lambda^{-1}).$$

An integration by parts for each integral in (31) shows that (31) is $O(\lambda^{-2}T)$. Arguments similar to those used at the end of the proof of (I) complete the proof of (I′).

6. Proof of (I″). Another integration by parts shows that (32) is

$$(1/4\nu_k)\{q'(t_{k+1})-q'(t_k)-\int_{t_k}^{t_{k+1}} q''\cos 2\nu_k^{\frac12}(t-t_k)\,dt\}.$$

Hence (30) is the sum of

$$(33)\quad -\tfrac12\lambda^{-\frac12}\{(1/4\lambda)\sum_{k=1}^{n}(q'(t_{k+1})-q'(t_k)) \\ -\sum_{k=1}^{n}(q'(t_{k+1})-q'(t_k))(1/4\lambda-1/4\nu_k)\}$$

and

$$(34)\quad O(\lambda^{-3/2})\sum_{k=1}^{n}\Big|\int_{t_k}^{t_{k+1}} q''\cos 2\nu_k^{\frac12}(t-t_k)\,dt\Big| = O(\lambda^{-3/2})\int_0^T \omega^{**}(t,\pi/\lambda^{\frac12})\,dt.$$

The first term of (33) is $O(\lambda^{-3/2})\,|\,q'(t_{n+2})-q'(t_1)\,|=O(\lambda^{-3/2})$, as $T\to\infty$, since $q'(t)$ is bounded. The second is $O(n\lambda^{-5/2})$ since $|\,q'(t_{k+1})-q'(t_k)\,| = O(1/\lambda^{\frac12})$ and $|\,\nu_k-\lambda\,| = O(1)$. Also, $n\lambda^{-\frac12}=O(T)$, by (13), as $T\to\infty$, for a fixed large λ.

Hence, $\theta(T,\lambda)$ is the sum of (29), (34) and terms of the type $O(\lambda^{-2}T)$, as $T\to\infty$. The proof of (I″) can now be completed as above.

7. Examples. The examples mentioned in Section 2 can be obtained by the consideration of periodic functions $q(t)$. If $q(t)$ is periodic, then the essential spectrum S' of (1) is the closure of the set of λ-values corresponding to which all solutions of (1) are bounded ([12]; cf. also [8]). Hence, if $\lambda=\lambda_k$ or λ^k, then (1) possesses a periodic or half-periodic solution.

If $q(t)$ is even and of period 2π, it is easy to see that the numbers $\lambda_1,\lambda_2,\cdots$ and $\lambda^0,\lambda^1,\cdots$ can be characterized as follows: Let $\alpha_0,\alpha_1,\cdots$ and $\beta_0,\beta_1,\cdots$ denote the eigenvalues belonging to the Sturm-Liouville problem of (1) and the boundary conditions $x(0)=x(\pi)=0$ and $x'(0)=x'(\pi)=0$, respectively. Then $\beta_0=\lambda^0$ and the numbers λ_k, λ^k for $k\geqq 1$ are, up to a permutation, the numbers α_{k-1}, β_k. Hence

$$\lambda^k - \lambda_k = |\alpha_{k-1} - \beta_k| \tag{35}$$

for $k = 1, 2, \cdots$.

On the other hand, if $q(t)$ has the Fourier series

$$q(t) \sim \sum_{n=1}^{\infty} a_{2n} \cos 2nt, \qquad (a_0 = 0), \tag{36}$$

then $\alpha_{k-1} = k^2 + \tfrac{1}{2}a_{2k} + o(1/k)$ and $\beta_k = k^2 - \tfrac{1}{2}a_{2k} + o(1/k)$, as $k \to \infty$; cf. [10], p. 322; also, [1], p. 11. Consequently,

$$\lambda^k - \lambda_k = |a_{2k}| + o(1/k) \text{ and } \lambda_k = k^2 + o(1), \tag{37}$$

as $k \to \infty$.

If $\psi(\lambda)$ is an arbitrary positive function which tends monotonously to 0, as $\lambda \to \infty$, then there exist continuous functions $q(t)$ for which $a_{2k} = O(\psi(k^2 - 1))$ fails to hold as $k \to \infty$. For such functions q, it follows from (31) that $\lambda^k - \lambda_k = O(\psi(\lambda_k))$ fails to hold as $k \to \infty$.

In order to obtain the second type of example mentioned in Section 2, it is sufficient to remark that there exist continuous functions q, of bounded variation over a period, for which $a_{2k} = O(1/k)$ but $a_{2k} \neq o(1/k)$, as $k \to \infty$; [2]. For such functions q, it follows from (37) that

$$\lambda^k - \lambda_k = O(1/\lambda_k^{\frac{1}{2}}) \text{ but } \lambda^k - \lambda_k \neq o(1/\lambda_k^{\frac{1}{2}})$$

as $k \to \infty$. This shows that, if (6) is assumed, the O in (5) cannot in general be replaced by o.

8. Proof of (II). Corresponding to any solution $x = x(t) = x(t, \lambda) \not\equiv 0$ of (1), let $h(t)$ be defined by

$$h(t) = \lambda x^2(t) + x'^2(t). \tag{38}$$

It follows from (1) that $h'(t) = -2qxx'$, so that

$$h'(t)/h(t) = -2q\{xx'/(\lambda x^2 + x'^2)\}.$$

Consequently,

$$\lambda x^2 + x'^2 = \text{const. } e^{At}, \tag{39}$$

where

$$A = A(t, \lambda) = -t^{-1} \int_0^t q\{2xx'/(\lambda x^2 + x'^2)\} ds. \tag{40}$$

Let $(0 \leqq)\ t_1 < t_2 < \cdots$ be the successive zeros of x. Since the factor $\{\cdots\}$ of q in the integral in (40) is homogeneous in x, x', the solution x can be replaced by any solution y of (1) vanishing at $t = t_k$.

From (23) and (25), it is seen that

$$2yy'/(\lambda y^2+y'^2)=\lambda^{-\frac{1}{2}}\sin 2\nu_k^{\frac{1}{2}}(s-t_k)+O(\lambda^{-3/2}), \tag{41}$$

as $\lambda\to\infty$, uniformly for $t_k\leqq s\leqq t_{k+1}$. Hence, for every fixed large λ,

$$A(t,\lambda)=-(1/t\lambda^{\frac{1}{2}})\sum_{k=1}^{n}\int_{t_k}^{t_{k+1}}q(s)\sin 2\nu_k^{\frac{1}{2}}(s-t_k)\,ds+O(\lambda^{-3/2})$$

as $t\to\infty$, where $n=n(t)$ is determined by $t_n\leqq t<t_{n+1}$. As in the proof of (I), it follows that

$$|A(t,\lambda)|\leqq(1/t\lambda^{\frac{1}{2}})\int_0^t\omega(s,\pi/\lambda^{\frac{1}{2}})\,d\lambda+O(\lambda^{-3/2}).$$

In view of the definition of $K(\lambda)$, this completes the proof of (9).

9. Proof of (II′) and (II″). Under the conditions of (II′), the arguments leading to (28) hold. Similar arguments show that $2yy'/(\lambda y^2+y'^2)$, up to a correction term $O(\lambda^{-2})$, is

$$(\nu_k^{\frac{1}{2}}/\lambda)\,(\sin 2\nu_k^{\frac{1}{2}}(s-t_k)-q(s)\lambda^{-1}\cos^2\nu_k^{\frac{1}{2}}(s-t_k)\sin 2\nu_k^{\frac{1}{2}}(s-t_k)),$$

as $\lambda\to\infty$, uniformly for $t_k\leqq s\leqq t_{k+1}$. Consequently, $A(t,\lambda)$, up to a correction term $O(\lambda^{-2})$, is the sum of the three expressions

$$-\sum_{k=1}^{n}(\nu_k^{\frac{1}{2}}/\lambda t)\int_{t_k}^{t_{k+1}}q(s)\sin 2\nu_k^{\frac{1}{2}}(s-t_k)\,ds, \tag{42}$$

$$\sum_{k=1}^{n}(\nu_k^{\frac{1}{2}}/2\lambda^2 t)\int_{t_k}^{t_{k+1}}q^2(s)\sin 2\nu_k^{\frac{1}{2}}(s-t_k)\,ds,$$

$$\sum_{k=1}^{n}(\nu_k^{\frac{1}{2}}/4\lambda^2 t)\int_{t_k}^{t_{k+1}}q^2(s)\sin 4\nu_k^{\frac{1}{2}}(s-t_k)\,ds.$$

Integrations by parts show that (42) is

$$(1/2\lambda t)\sum_{k=1}^{n}\{q(t_{k+1})-q(t_k)-\int_{t_k}^{t_{k+1}}q'(s)\cos 2\nu_k^{\frac{1}{2}}(s-t_k)\,ds\},$$

which is

$$O(1/\lambda t)-(1/2\lambda t)\sum_{k=1}^{n}\int_{t_k}^{t_{k+1}}q'(s)\cos 2\nu_k^{\frac{1}{2}}(s-t_k)\,ds, \tag{43}$$

since $q(t)$ is bounded. Integrations by parts also show that both expressions in the formula lines following (42) are $O(1/\lambda^2)$, as $t \to \infty$, for every fixed large λ. Hence,

$$|A(t,\lambda)| \leqq (1/2\lambda t) \int_0^t \omega^*(s, \pi/\lambda^{\frac{1}{2}})\,ds + O(1/\lambda^2),$$

as $t \to \infty$, for large λ. The relation (10) follows.

Under the conditions of (II″), another integration by parts can be applied to each of the integrals in (43), thereby yielding the inequality

$$|A(t,\lambda)| \leqq O(1/\lambda^{3/2}t) \int_0^t \omega^{**}(s, \pi/\lambda^{\frac{1}{2}})\,ds + O(1/\lambda^2),$$

thus completing the proof of (II″).

Appendix.

It was shown in [7], p. 137, that if q satisfies (13) and if $|\lambda| > M$, then a necessary condition for λ to be in the complement of S', is that there exist a number $l > 0$ and a pair of solutions $x_1(t)$, $x_2(t)$ of (1) satisfying (8). It is the purpose of this Appendix to show that this necessary condition is not sufficient. To this end, the example in [13], p. 269, will be refined.

Let the seqeuence of numbers $m_0 < m_1 < \cdots$ be defined as follows:

$m_0 = 0$, $m_{2n+1} = m_{2n} + 2(n+1)\pi$ and $m_{2n+2} = m_{2n+1} + 2(n+1)\pi$ for $n \geqq 0$.

This sequence has the properties that m_n is an even integral multiple of π and that

$$t^{-1} \int_0^t \epsilon(s)\,ds \geqq 1 \text{ if } t > 0, \tag{44}$$

and $\epsilon(s)$ is 2 or 0 according as $m_{2n} \leqq s < m_{2n+1}$ or $m_{2n+1} \leqq s < m_{2n+2}$ for $n = 0, 1, \cdots$.

In terms of $\epsilon(s)$, the sequence $m_0, m_1, \cdots$ and a number $\alpha > 0$, define a function $\chi(s)$, where $0 \leqq s < \infty$, by placing

$$\chi(s) = \alpha\epsilon(s)(1 - \cos s). \tag{45}$$

It is easily seen that $\chi(s)$ possesses a continuous derivative (even at $s = m_0, m_1, \cdots$). It follows from [13], p. 268, that if

$$q(t) = -\epsilon(t)\{2\alpha^2(1-\cos t)^2\cos^2 t - 3\alpha\sin t + 2\alpha\sin 2t\} \tag{46}$$

and $\lambda = 1$, then (1) has

$$(47) \qquad x = x(t) = \exp\left(\int_0^t \chi(s) \cos s \, ds\right) \cos t$$

as a solution.

From (45), it is seen that

$$\int_0^t \chi(s) \cos s \, ds = -\alpha \int_0^t \epsilon(s) \left(\tfrac{1}{2} + \tfrac{1}{2} \cos 2s - \cos s\right) ds;$$

and from (44), that

$$\int_0^t \chi(s) \cos s \, ds \leqq -\tfrac{1}{2}\alpha t + O(1)$$

as $t \to \infty$. Hence, $x^2(t) \leqq \text{const.}\, e^{-\alpha t}$ and $x'^2(t) \leqq \text{Const.}\, e^{-\alpha t}$, by (47). Thus, if $x_1(t)$ is defined as a suitable multiple of (47), the first inequality in (8) is satisfied by $l = \alpha > 0$. If $x_2(t)$ is any solution linearly independent of $x_1(t)$, the second inequality of (8) holds with $l = \alpha$; cf., e. g., [7], p. 138.

If $\alpha > 0$ is sufficiently small, it follows from (46) that $|q| \leqq M < 1$. Hence, in order to complete the proof, it remains to show that $\lambda = 1$ is in S' in the case (46). Actually, it will be shown that S' contains the half-line $0 \leqq \lambda < \infty$, by proving that (12) holds for every $\lambda > 0$ and $\Delta > 0$.

The Sturm comparison theorem implies that if $T > m_{2n}$, then $N(T, \lambda + \Delta) - N(T, \lambda)$ is not less than $[N(m_{2n}, \lambda + \Delta) - N(m_{2n-1}, \lambda + \Delta)] - [N(m_{2n}, \lambda) - N(m_{2n-1}, \lambda)] - 2$. On the other hand, since $q(t) = 0$ for $m_{2n-1} \leqq t < m_{2n}$, it follows from (1) that

$$N(m_{2n}, \lambda + \Delta) - N(m_{2n-1}, \lambda + \Delta) \geqq (\lambda + \Delta)^{\frac{1}{2}} (m_{2n} - m_{2n-1})/\pi - 2$$

and

$$N(m_{2n}, \lambda) - N(m_{2n}, \lambda) \leqq \lambda^{\frac{1}{2}} (m_{2n} - m_{2n-1})/\pi + 2.$$

It is seen, from the definition of the sequence $m_0, m_1, \cdots$, that if $T > m_{2n}$, then

$$N(T, \lambda + \Delta) - N(T, \lambda) \geqq ((\lambda + \Delta)^{\frac{1}{2}} - \lambda^{\frac{1}{2}}) 2n - 6.$$

This completes the proof.

A slight modification of the above example leads to the following: The function

$$(48) \qquad x = x(s) = \exp\left(-\int_0^t \chi(s) \, ds\right)$$

is a solution of (1) if $\lambda = 0$ and

$$q(t) = -\epsilon(t)(2\alpha^2 \sin t - \alpha \cos t). \tag{49}$$

It is easily seen that (48) satisfies $x^2 + x'^2 = O(e^{-2\alpha t})$, as $t \to \infty$. Hence, there exist linearly independent solutions satisfying (8) with $l = 2\alpha > 0$. Furthermore, in the case (49), the differential equation (1) is oscillatory for $\lambda > 0$, since $q(t) = 0$ for $m_{2n-1} \leqq s < m_{2n-1} + 2n\pi$. Since (48) has no zeros, it follows that $\lambda = 0$ is the least point of S'; [5]. This is in contrast to (47) which belongs to an interior point of S' in the case (46).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] G. Borg, "Eine Umkehrung der Sturm-Liouvillesche Eigenwertaufgabe. Bestimmung der Differentialgleichung durch die Eigenwerte," *Acta Mathematica*, vol. 78 (1946), pp. 319-334.

[2] F. Riesz, "Ueber die Fourierkoeffizienten einer stetigen Funktion von beschränkter Schwankung," *Mathematische Zeitschrift*, vol. 2 (1918), pp. 312-315.

[3] P. Hartman, "On the spectra of slightly disturbed linear oscillators," *American Journal of Mathematics*, vol. 71 (1949), pp. 71-79.

[4] ———, "A characterization of the spectra of one-dimensional wave equations," *ibid.*, vol. 71 (1949), pp. 915-920.

[5] ——— and C. R. Putnam, "The least cluster point of the spectrum of boundary value problems," *ibid.*, vol. 70 (1948), pp. 849-855.

[6] C. R. Putnam, "The cluster spectra of bounded potentials," *ibid.*, vol. 71 (1949), pp. 612-620.

[7] ———, "On isolated eigenfunctions associated with bounded potentials," *ibid.*, vol. 72 (1950), pp. 135-147.

[8] S. Wallach, "The spectra of periodic potentials," *ibid.*, vol. 70 (1948), pp. 842-848.

[9] H. Weyl, "Ueber gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkurlicher Funktionen," *Mathematische Annalen*, vol. 68 (1910), pp. 222-269.

[10] ———, "Die Gibbs'sche Erscheinung in der Theorie der Kugelfunktionen," *Rendiconti del Circolo Matematico di Palermo*, vol. 29 (1910), pp. 308-323.

[11] A. Wintner, "The adiabatic linear oscillator," *American Journal of Mathematics*, vol. 68 (1946), pp. 385-397.

[12] ———, "Stability and spectrum in the wave mechanics of lattices," *Physical Review*, vol. 72 (1947), pp. 81-82.

[13] ———, "Asymptotic integration of the adiabatic oscillator," *American Journal of Mathematics*, vol. 69 (1947), pp. 251-272.

[14] ———, "On the smallness of isolated eigenfunctions," *ibid.*, vol. 71 (1949), pp. 603-611.

ON NON-UNIFORMLY QUASI-ANALYTIC FUNCTIONS.*

By I. I. Hirschman, Jr.

---

1. Introduction. In previous papers, see [1] and [2], various results have been established concerning functions which belong to quasi-analytic classes $-\infty < x < \infty$. For example, it has been shown that if

$$(1) \qquad | f^{(n)}(x) | \leqq Ak^n n! \, [\log(n+e)]^n, \qquad -\infty < x < \infty \, ; n = 0, 1, \cdots,$$

and if

$$(2) \qquad f(x) = O(\exp[-\exp\exp\{\pi \ (1+\epsilon)/2\}kx]), \qquad x \to +\infty,$$

where $\epsilon > 0$, then $f(x) \equiv 0$. It is natural to inquire whether similar conclusions can be drawn when the inequalities (1) are replaced by the more general inequalities

$$(3) \quad | f^{(n)}(x) | \leqq A[k(x)]^n n! [\log(n+e)]^n, \quad -\infty < x < \infty, n = 0, 1, \cdots.$$

In the present paper we shall obtain some results of this kind. We shall, for example, show that if $f(x)$ satisfies the inequalities (3) where $k'(x)[k(x)]^{-2} = o(1)$ as $x \to +\infty$, and if

$$(4) \qquad f(x) = O(\exp[-\exp\exp \tfrac{1}{2}\pi(1+\epsilon) \int_0^x k(t)\,dt]), \qquad x \to +\infty$$

with $\epsilon > 0$, then $f(x) \equiv 0$. This result reduces to the previous one when $k(x)$ is constant. The method of demonstration, which is very simple, depends upon conformal mapping of curvilinear strips.

2. Preliminary results. We shall take $z = x + iy$ and $w = u +$ [illegible] throughout.

Theorem 2a. *Suppose that $f(x)$ is infinitely differentiable in a neighborhood of $x = a$, $\phi(z)$ is analytic for $| z - \alpha | < \rho$, $\phi(\alpha) = a$, and $\phi(x)$ is real for $| x - \alpha | < \rho$. If*

1. $| f^{(n)}(a) | \leqq Ak^n M_n$, $n = 0, 1, \cdots$;
2. $M_n q^n / n!$ *is non decreasing for a fixed* $q > 0$;
3. $| \phi(z) - a | \leqq r$ *for* $| z - \alpha | \leqq \rho$;
4. $l = q\rho^{-1} \operatorname{Max}[1, krq^{-1}]$,

then for $F(x) = f[\phi(x)]$ *we have* $| F^{(n)}(\alpha) | \leqq nAl^n M_n$, $n = 1, 2, \cdots$.

---

* Received April 26, 1950

Formally,

$$f(x) = \sum_{j=0}^{\infty} [f^{(j)}(a)/j!](x-a)^j, \quad F(x) = \sum_{j=0}^{\infty} [f^{(j)}(a)/j!][\phi(x)-a]^j,$$

and thus

$$(1) \qquad F^{(n)}(\alpha) = \sum_{j=1}^{n} [f^{(j)}(a)/j!](d/dx)^n[\phi(x)-a]^j{}_{x=a}, \qquad n \geqq 1,$$

(the terms corresponding to $j = 0$ and $j > n$ are zero). It is easily seen that equation (1) is valid generally. Cauchy's formula yields that

$$(d/dx)^n[\phi(x)-a]^j{}_{x=a} = (2\pi i)^{-1} n! \int_{|z-a|=\rho} \{[\phi(z)-a]^j/[z-\alpha]^{n+1}\} dz,$$

$$(2) \qquad |(d/dx)^n[\phi(x)-a]^j{}_{x=a}| \leqq n!r^j\rho^{-n}.$$

Inserting inequality (2) in equation (1) we obtain

$$|F^{(n)}(\alpha)| \leqq \sum_{j=1}^{n} [M_j q^j/j!][An!\rho^{-n}k^j q^{-j} r^j],$$

$\leqq nAl^nM_n$, as desired.

THEOREM 2b. *If*

1. $\theta(x)$ *is defined for* $-\infty < x < \infty$; $\theta(x) > 0$; $\theta'(x) = o(1)$ *as* $x \to +\infty$;
2. S_z *is the strip* $|y| < \theta(x)$, $--\infty\ x < \infty$;
3. S_w *is the strip* $|v| < 1$, $-\infty < u < \infty$;
4. $\psi(z)$ *maps* S_z *conformally onto* S_w, $\psi(0) = 0$, $\psi'(0) > 0$, *and* $\phi(w)$ *is inverse to* $\psi(z)$,

[illegible]en for $0 < \lambda < 1$ *we have*

A. $\limsup\limits_{x_1\to+\infty} \max\limits_{|z-x_1|\leqq\lambda\theta(x_1)} |\psi(z)-\psi(x_1)| = \lambda$,

B. $\limsup\limits_{u_1\to+\infty} \max\limits_{|w-u_1|\leqq\lambda} |\phi(w)-\phi(u_1)|/\theta[\phi(u_1)] = \lambda$,

C. $\psi(x) \sim \int_0^x [\theta(t)]^{-1} dt, \qquad x \to +\infty.$

Let $\eta(x_1, x) = \theta[x_1 + \theta(x_1)x]/\theta(x_1)$ and let $S(x_1)$ be the strip $|y| < \eta(x_1, x)$, $-\infty < x < \infty$. It may then be verified that the function $\psi(x_1, z) = \psi[\theta(x_1)z + x_1] - \psi(x_1)$ maps $S(x,)$ onto S_w conformally and that $\psi(x_1, 0) = 0$, $\psi'(x_1, 0) > 0$. It follows from assumption 1 that $\lim\limits_{x_1\to+\infty} \eta(x_1, x) = 1$ uniformly for x in any finite interval. By a well known theorem on convergent sequences of domains, see [3; § 123], it follows that

$$\lim \psi[\theta(x_1)z + x_1] - \psi(x_1) = z, \qquad (x_1 \to \infty), \tag{3}$$

uniformly for z in any compact subset of the strip $|y| < 1$, $-\infty < x < \infty$. Conclusions A. and B. are immediate consequences of this fact. To demonstrate conclusion C. we choose a sequence $\{x_n\}_0^\infty$, $0 = x_0 < x_1 < x_2 < \cdots$, such that $r' < (x_{n+1} - x_n)/\theta(x_n) < r''$, $n = 0, 1, \cdots$, for some numbers $0 < r' < r'' < \infty$. It follows from equation (3) that

$$\psi(x_{n+1}) - \psi(x_n) \sim (x_{n+1} - x_n)/\theta(x_n).$$

From assumption 1 we may deduce that $(x_{n+1} - x_n)/\theta(x_n) \sim \int_{x_n}^{x_{n+1}} [\theta(t)]^{-1} dt$. Hence summing $\psi(x_n) \sim \int_0^{x_n} [\theta(t)]^{-1} dt$. Since any value of x lies between two values of the sequence $\{x_n\}_0^\infty$ and since $\int_{x_n}^{x_{n+1}} [\theta(t)]^{-1} dt = O(1)$ as $n \to \infty$, our desired result follows.

LEMMA 2c. *We have*

$$\operatorname*{Max}_{|z-x_1| \leqq 1, -\infty < x_1 < \infty} |[z^2 + c^2]^{1/2} - [x_1^2 + c^2]^{1/2}| = 1 + R(c),$$

where $R(c) = o(1)$ *as* $c \to +\infty$.

The proof of this simple result is left to the reader.

3. Applications. THEOREM 3a. *Let*

1. $\{M_n\}_0^\infty$ *be a logarithmically convex sequence of non-negative constants such that* $\limsup [n!/M_n]^{1/n} < \infty$;
2. $T(r) = \operatorname*{Max}_{n \geqq 0} r^n/M_n$;
3. $\int_1^\infty [\log T(r)] r^{-2} dr = \infty$, $t = (2/\pi) \int_1^{\mathfrak{H}(t)} [\log t(r)] r^{-2} dr$;
4. $k(x) > 0$, $0 \leqq x \leqq \infty$, $k'(x)[k(x)]^{-2} = o(1)$, $x \to +\infty$;
5. $|f^{(n)}(x)| \leqq AM_n[k(x)]^n$, $0 \leqq x < \infty$; $n = 0, 1, \cdots$.

Then if $f(x) = O(\exp\{-\mathfrak{H}[(1+\epsilon) \int_0^x k(t)\,dt]\})$ *as* $x \to +\infty$, *where* $\epsilon > 0$, *we have* $f(x) \equiv 0$.

By logarithmic convexity we have that $M_{n-1} \leqq M_n^{(n-1)/n} M_0^{1/n}$, and hence that $[M_{n-1}/(n-1)!][M_n/n!]^{-1} \leqq nM_0^{1/n} M_n^{-1/n}$. It follows from assumption 1 that $\limsup_{n \to \infty} nM_0^{1/n} M_n^{-1/n} < \infty$, and thus that

$$q = \operatorname*{Max}_{n \geqq 1} [M_{n-1}/(n-1)!][M_n/n!]^{-1}$$

is finite. Clearly, $M_n q^n/n!$ is non-decreasing.

We define $\theta(x) = q/k(0)$ for $-\infty < x \leqq 0$ and $\theta(x) = q/k(x)$ for $0 < x < \infty$. Because of assumption 4. we have $\theta'(x) = o(1)$ as $x \to +\infty$. Let S_z, S_w, and $\psi(z)$ and $\phi(w)$ be defined as in Theorem 2c. Consider the function $g(u) = f[\phi(u)]$. By Theorem 2a and conclusion B of Theorem 2b we see that, given $\delta > 0$, there exists a u_0 such that $| g^{(n)}(u) | \leqq A_0 q^n (1+\delta)^n M_n$ for $u \geqq w_0$, $n = 0, 1, \cdots$. By Theorem 2a and Lemma 2c, given $\delta_1 > \delta$, there exists a value of c such that if $h(t) = g[(t^2+c^2)^{\frac{1}{2}}]$, then

$$| h^{(n)}(t) | \leqq A_1 q^n (1+\delta_1)^n M_n, \qquad -\infty < t < \infty \,; n = 0, 1, \cdots.$$

We have $g[\psi(x)] = f(x)$, and so, by assumption,

$$g[\psi(x)] = O(\exp\{-\mathfrak{H}[(1+\epsilon)\int_0^x k(t)\,dt]\})$$

as $x \to +\infty$. By conclusion C of Theorem 2c we have for $0 < \epsilon_0 < \epsilon$ and x large that

$$(1+\epsilon)\int_0^x k(t)\,dt > (1+\epsilon_0)q\psi(x).$$

It follows that

$$g(x) = O(\exp\{-\mathfrak{H}[(1+\epsilon_0)qx]\})$$

as $x \to +\infty$. If $0 < \epsilon_1 < \epsilon_0$, then

$$h(t) = O(\exp\{-\mathfrak{H}[(1+\epsilon_1)qt]\})$$

as $t \to +\infty$. We may suppose that $(1+\delta_1)/(1+\epsilon_1) < 1$. It then follows from Theorem 4 of [1] that $h(t) \equiv 0$, hence $f(x) \equiv 0$.

THEOREM 3b. *Let*

1. $M_n = n!\,\nu(n)^n$, *where* $\nu'(x) \geqq 0$, *and* $x\nu'(x)/\nu(x) = o(1)$, $x \to \infty$;
2. $T(r) = \operatorname*{Max}_{n \geqq 0} r^n/M_n$;
3. $\int_1^\infty \log T(r) r^{-2}\,dr = \infty$, $H(t) = (2/\pi)\int_1^t \log T(r) r^{-2}\,dr$;
4. $k(x) > 0$ $(0 \leqq x < \infty)$, $k'(x)[k(x)]^{-2} = o(1)$, $x \to +\infty$;
5. $| f^{(n)}(x) | \leqq A[k(x)]^n M_n$, $0 \leqq x < \infty \,; n = 0, 1, \cdots$;
6. $Z(u)$ *be the number of zeros of* $f(x)$ *for* $0 \leqq x \leqq u$, *counted according to their multiplicities.*

Then if $f(x) \not\equiv 0$, *we have*

$$\limsup_{x \to +\infty} H[Z(x)] \Big/ \int_0^x k(t)\,dt \leqq 1.$$

The demonstration of this theorem is just like that of Theorem 3a, except that we use Theorem 4b of [2] in place of Theorem 4 of [1].

It is possible to show that the conclusion of Theorem 3a is false for $\epsilon < 0$. One can employ for this purpose the counter-example given in [1], together with Theorem 2a and conclusions A. and C. of Theorem 3b. Similarly, Theorem 3b cannot be greatly improved.

WASHINGTON UNIVERSITY.

BIBLIOGRAPHY.

[1] I. I. Hirschman, Jr., "On the behaviour of Fourier transforms at infinity and on quasi-analytic classes of functions," *American Journal of Mathematics*, vol. 72 (1950), pp. 200-213.

[2] ———, "On the distributions of the zeros of functions belonging to certain quasi-analytic classes," *ibid.*, vol. 72 (1950), pp. 396-406.

[3] C. Carathéodory, *Conformal Representation*, Cambridge, 1941.